Bruce Hartpence
Übersetzung von Peter Klicman

o'reillys basics

Praxiskurs Routing & Switching

- Routing- und Switching-Strategien, statische und dynamische Topologien
- (Rapid) Spanning Tree Protocol, VLAN Trunking Protocol
- Routing Information Protocol, Open Shortest Path First

O'REILLY®

Praxiskurs Routing und Switching

Bruce Hartpence

Deutsche Übersetzung von Peter Klicman

O'REILLY®

Beijing · Cambridge · Farnham · Köln · Sebastopol · Tokyo

Kommentare und Fragen können Sie gerne an uns richten:
O'Reilly Verlag GmbH & Co. KG
Balthasarstr. 81
50670 Köln
E-Mail: kommentar@oreilly.de

1. Auflage 2012

Die Originalausgabe erschien 2011 unter dem Titel
Packet Guide to Routing and Switching bei O'Reilly Media, Inc.

Die Darstellung eines Grünen Heupferds im Zusammenhang mit dem Thema Netzwerke ist ein Warenzeichen des O'Reilly Verlags GmbH & Co. KG.

Bibliografische Information Der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über *http://dnb.d-nb.de* abrufbar.

Übersetzung: Peter Klicman, Köln
Lektorat: Imke Hirschmann, Köln
Korrektorat: Friederike Daenecke, Zülpich
Produktion: Andrea Miß, Köln
Umschlaggestaltung: Micheal Oreal, Köln
Satz: Reemers Publishing Services GmbH, Krefeld, www.reemers.de
Druck: Druckerei Kösel, Krugzell, www.koeselbuch.de

ISBN: 978-3-86899-185-7

Dieses Buch ist auf 100% chlorfrei gebleichtem Papier gedruckt.

Für Christina, Brooke, Nick und Sydney –
meinen ewigen Dank für die Liebe und das Lachen,
das noch kommt.

Inhalt

Vorwort

Lange Zeit war ich damit zufrieden, Ethernet-Netzwerke aufzubauen, mit Switches zu arbeiten, und irgendwann zu 802.11 zu wechseln. Es dauerte eine Weile, doch schließlich erkannte ich, dass die Welt verbundener Netzwerke nicht allein mit der zweiten Schicht zu bewältigen ist. Und wenn man dann von den Schicht-2-Broadcast-Domains über den Tellerrand schaut, entdeckt man die Wunder virtueller lokaler Netzwerke und Trunks. Ich wurde zu einer Art »ganzheitlichem« Netzwerk-Typ. Mit meiner eigenen Weiterentwicklung bewegt sich auch dieses Buch hin zu den nächsten Ebenen und Ideen.

Wenn Sie *Praxiskurs Netzwerkgrundlagen* (O'Reilly) gelesen haben, haben Sie eine Vorstellung davon, welche Art der Kommunikation (ARP, ICMP, IP, Ethernet) in einem Netzwerk vorkommt, und zwar unabhängig von Betriebssystem und Netzwerkausrüster. Dieses Buch macht nun den Schritt in Richtung fortgeschrittene Link- und Internetwork-Layer-Protokolle, die es Ihnen ermöglichen, sich mit Internetworking und großen Topologien zu beschäftigen.

Wie im ersten Buch nimmt sich jedes Kapitel ein bestimmtes Protokoll oder eine Idee vor und erläutert dessen Struktur und Funktionsweise. Diese Betrachtung wird durch umfangreiche Paket-Captures unterstützt. An den Dingen in diesem Buch ist nichts Theoretisches: Die für jedes Kapitel gewählten Topologien wurden in einem Labor aufgebaut, während die einzelnen Kapitel Form annahmen.

Und genau wie im ersten Buch ist alles, was Sie hier sehen, Teil jedes Netzwerks, mit dem Sie arbeiten. Die hier vorgestellten Praktiken, Ideen und Protokolle werden Ihnen also in den nächsten Jahren hilfreich zur Seite stehen. Ich werde auch immer wieder mit Netzwerktabellen für das Routing (Host und Router), Quelladressen und ARP-Tabellen arbeiten.

Zuletzt haben viele Netzwerk-Profis den IPv6-Tag erlebt oder ihm zumindest eine gewisse Aufmerksamkeit gewidmet. Doch die Ergebnisse waren größtenteils unbefriedigend. Verschiedene Herausforderungen – etwa der korrekte Betrieb von 6-nach-4-Tunneln, Filtern für bestimmte IPv6-Nachrichten und das Fehlen von Sicherheitsfeatures – lassen vermuten, dass uns IPv4 noch einige Zeit erhalten bleiben wird. Gleichwohl sprechen viele Kapitel IPv6 an, einschließlich der grundlegenden Konfiguration und einem Vergleich zum IPv4-Betrieb.

Jedes Kapitel enthält eine Reihe abschließender Fragen, um Sie an die Schlüsselideen zu erinnern. Eine Reihe von (einfachen bis fortgeschrittenen) Laborexperimenten ist ebenfalls enthalten. Die praktischen Übungen wurden so gestaltet, dass Sie sie mithilfe des jeweiligen Kapitels durchführen und die Ideen gleich in die Tat umsetzen können.

Ich hoffe, dass Sie Spaß an diesem Buch haben und dass es Ihnen bei Ihrer Arbeit hilft.

Leserkreis

Da dieses Buch sowohl grundlegende Dinge als auch fortgeschrittene Ideen erläutert, eignet es sich sowohl für den Einsteiger als auch für den Profi, der eine kleine Auffrischung braucht. Egal ob Sie mit kleinen Netzwerken arbeiten oder große Netzwerke miteinander verbinden, die hier vorgestellten Prinzipien sind immer gleich.

Dieses Buch ist als Ergänzung zum *Praxiskurs Netzwerkgrundlagen* gedacht. Beide Bücher sind eigenständig, doch dieses Buch geht davon aus, dass Sie die Konzepte und Protokolle verstehen, die im anderen Buch erläutert wurden, einschließlich ARP, ICMP, IP, Geräten, Ethernet und Maskierung. Gelegentlich gehe ich auf diese Dinge ein, doch das ist nur selten der Fall.

Inhalt des Buches

Kapitel 1, Routing- und Switchingstrategien

Dieses Kapitel ist der Kitt für dieses Buch. Es behandelt die ganzheitliche Natur der im Netzwerk getroffenen Weiterleitungs-(Forwarding-)Entscheidungen und stellt viele Konzepte vor, die die Grundlage späterer Kapitel bilden. Dieses Kapitel behandelt Schlüsselkonzepte wie die Klassifikation von Protokollen, den Vergleich von statischen gegenüber dynamischen

Topologien und die Gründe für die Installation einer bestimmten Route.

Kapitel 2, Host-Routing
: Dieses Kapitel macht da weiter, wo die Betrachtung von Masken in Kapitel 1 aufgehört hat. Hosts verhalten sich in vielerlei Hinsicht wie Router, und das trifft auch auf die Verarbeitung der Routingtabelle zu. Dieses Kapitel zeigt, wie eine Host-Routingtabelle verarbeitet wird und wie der Traffic seine Reise durch das Netzwerk beginnt. Die Kommunikation beim Durchlaufen von Routern wird ebenfalls untersucht, wobei wir der Adressierung und der Konstruktion von Frames besondere Aufmerksamkeit widmen.

Kapitel 3, Spanning Tree und Rapid Spanning Tree
: Schleifen sind für jedes Ethernet-Netzwerk problematisch. Das Spanning Tree Protocol ist integraler Bestandteil jedes Netzwerks, das Switches nutzt und schützt die Topologie vor solchen Schleifen. Es kann sich auch auf die Performance Ihres Netzwerks auswirken und Bandbreite verbrauchen. Dieses Kapitel behandelt das Spanning Tree Protocol und das schnellere Rapid Spanning Tree Protocol.

Kapitel 4, VLANs und Trunking
: So gut Switches für moderne Kommunikationstopologien auch sind, sobald das Schicht-2-Netzwerk über eine bestimmte Größe hinauswächst, treten Flaschenhälse und Sicherheitsprobleme ganz von selbst auf. VLANs sind ein nützliches Werkzeug, um diese Probleme anzugehen. Kapitel 4 behandelt das Design und den Betrieb von VLANs sowie Abschnitte zu Trunking-(Bündelungs-)Protokollen, die es VLANs erlauben, sich über viele Switches auszudehnen.

Kapitel 5, Routing Information Protocol
: Eines der ersten Distanzvektorprotokolle, RIP, wird häufig als Basis genutzt, um das dynamische Routing zu verstehen. Doch RIP hat auch seinen Platz in kleinen, modernen Kommunikationsnetzwerken. Dieses Kapitel behandelt den Betrieb und die Struktur von RIP. Es diskutiert auch Verbesserungen des dynamischen Routings wie Split Horizon, Poisoning, Count-to-Infinity und getriggerte Updates.

Kapitel 6, Open Shortest Path First
: OSPF ist ein Link-State-Protokoll und wird gegenüber Protokollen wie RIP als überlegen betrachtet. Dieses Kapitel beschreibt den Betrieb von Link-State-Protokollen und erklärt, warum die Konvergenzzeiten besser sind als bei Distanzvekto-

ren. Protokollstruktur, Adressierung und Betrieb werden mithilfe von Paket-Captures erläutert.

Typografische Konventionen

In diesem Buch werden die folgenden typografischen Konventionen verwendet:

Kursivschrift
: wird für Menütitel, -optionen und -buttons sowie neue Begriffe, URLs, E-Mail-Adressen, Dateinamen, Dateiendungen, Pfadnamen, Verzeichnisse und Unix-Dienstprogramme verwendet.

`Nichtproportionalschrift`
: wird für Befehle, Optionen, Switches, Variablen, Attribute, Schlüssel, Funktionen, Typen, Klassen, Namensräume, Methoden, Module, Eigenschaften, Parameter, Werte, Objekte, Events, Eventhandler, XML-Tags, HTML-Tags, Makros, Dateiinhalte und Befehlsausgaben verwendet.

`Nichtproportionalschrift fett`
: wird für Befehle oder anderen Text verwendet, den Sie wortwörtlich eingeben müssen.

`Nichtproportionalschrift kursiv`
: wird für Text verwendet, der durch Benutzereingaben ersetzt werden muss.

Tipp Zeigt einen Tipp, eine Empfehlung oder einen allgemeinen Hinweis an.

Warnung Zeigt eine Warnung an.

Codebeispiele zu diesem Buch

Dieses Buch soll Ihnen bei der Arbeit helfen. Es ist grundsätzlich erlaubt, den Code dieses Buches in Ihren Programmen und der Dokumentation zu verwenden. Hierfür ist es nicht notwendig, uns um Erlaubnis zu fragen, es sei denn, es handelt sich um eine größere Menge Code. So ist es beim Schreiben eines Programms, das einige Codeschnipsel dieses Buches verwendet, nicht nötig, sich mit uns in Verbindung zu setzen; beim Verkauf oder Vertrieb einer CD-ROM mit Beispielen aus O'Reilly-Büchern dagegen schon. Das Beantworten einer Frage durch Zitieren von Beispielcode erfordert keine

Erlaubnis. Verwenden Sie einen erheblichen Teil des Beispielcodes aus diesem Buch in Ihrer Dokumentation, ist jedoch unsere Erlaubnis nötig.

Eine Quellenangabe ist zwar erwünscht, aber nicht unbedingt notwendig. Hierzu gehört in der Regel die Erwähnung von Titel, Autor, Verlag und ISBN. Zum Beispiel: »*Praxiskurs Routing und Switching* von Bruce Hartpence (O'Reilly). Copyright 2011 Bruce Hartpence, 978-3-86899-185-7«.

Falls Sie nicht sicher sind, ob die Nutzung der Codebeispiele über die hier erteilte Genehmigung hinausgeht, nehmen Sie bitte unter der Adresse *permissions@oreilly.com* Kontakt mit uns auf.

Danksagung

Dieses Buch folgt unmittelbar auf das erste, was bedeutet, dass die Mitglieder meiner Familie und die Kollegen in meiner Abteilung nun seit mehreren Monaten mit meinen schriftstellerischen Aktivitäten leben müssen. Wie nicht anders zu erwarten war, gab es überall Kabel, Hinweisschilder an Geräten, hohen Kaffeebedarf und allgemeine Verdrießlichkeit. Mein Dank gilt allen, die meine Schikanen ertragen haben.

Ich möchte mich bei den Leuten von O'Reilly bedanken, die das Schreiben beider Bücher zu einer tollen Erfahrung gemacht und einem neuen Autor auf die Beine geholfen haben.

Besonderer Dank gilt meinem Autor-Gewissen Jim Leone, der mir dabei half, nicht vom Weg abzukommen und mich vor zu vielen Pronomen beschützt hat. Ich hatte auch viel Hilfe von Jonathan Weissman, der nicht nur meine Liebe für alles Vernetzte teilt, sondern auch dafür sorgte, dass ich die richtige Reihenfolge einhielt und Redundanzen vermied.

KAPITEL 1

Routing- und Switchingstrategien

In diesem Kapitel:

- Switching: Traffic weiterleiten und filtern
- Routing: Wege finden
- IPv6
- Lektüre
- Zusammenfassung
- Fragen zum Kapitel
- Antworten
- Laborübungen

Das vorangegangene Buch dieser Serie, *Praxiskurs Netzwerkgrundlagen*, behandelte die IPv4-Protokolle, die Maskierung und die Geräte, die Teil jedes Netzwerks sind. Nun ist es an der Zeit, sich des Routings und Switchings des Netzwerks anzunehmen. Es gibt eine erstaunliche Zahl tabellenbasierter Entscheidungen, die getroffen werden müssen, damit ein einzelnes Paket durch ein einzelnes Netzwerk laufen kann, ganz zu schweigen von mehreren Netzwerken. Diese Entscheidungen sind nicht auf Router, Switches und Access Points beschränkt, sondern werden auf jedem Gerät getroffen, auch auf den Hosts. Während die Netzwerke aufgebaut und Geräte konfiguriert werden, um Pakete und Frames weiterzuleiten, müssen Netzwerkadministratoren kritische Entscheidungen treffen, die die Performance, Sicherheit und Optimierung beeinflussen.

Wendet man sich als Netzwerkadministrator fortgeschrittenen Ideen zu, muss man wissen, wie und warum Netzwerktabellen aufgebaut werden und in welchen Fällen Veränderungen von Hand vorteilhaft sind. Dieses Kapitel enthält Details zu Routing- und Switchingoperationen, aber auch zu Designelementen. Es wird vorausgesetzt, dass Sie die grundlegende Funktionsweise von Routern und Switches verstehen, sowie die üblichen Standardprotokolle wie Ethernet, Internet Protocol (IP), Address Resolution Protocol (ARP) und das Internet Control Message Protocol (ICMP) kennen.

Switching: Traffic weiterleiten und filtern

Die meisten Protokolle stehen von vornherein fest, d.h., dass Sie beim Aufbau eines Netzwerks in vielen Fällen gar keine Wahl haben. Es ist sehr wahrscheinlich, dass ein Netzwerk eine Mischung aus Ethernet- und 802.11-Knoten (Nodes) ist. Diese Knoten führen das Internet Protocol in der dritten Schicht des Transmission Con-

trol Protocol/Internet Protocol-(TCP/IP-)Netzwerkmodells (siehe Abbildung 1-1) aus. Die Anwendungen werden für TCP oder UDP (das User Datagram Protocol) ausgelegt.

Es gibt viele Switchingtypen: Packet, Circuit, Multilayer, Virtual Circuit, WAN (Wide Area Network), LAN (Local Area Network). Circuit Switching (oder Leitungsvermittlung) und Virtual Circuit Switching beziehen sich fast immer auf WAN- oder Telefontechniken und werden hier daher nicht weiter vertieft. Das Packet Switching betrifft üblicherweise Router und vielleicht einen WAN-Switch. Multilayer Switching ist eine Technik, die die Verarbeitung von IP-Paketen verbessern soll, doch die meisten Hersteller haben unterschiedliche Vorstellungen vom besten Ansatz. LAN-Switches werden häufig eingesetzt, ohne weiter darüber nachzudenken, wie das Multilayer Switching die Performance verbessern könnte. Tatsächlich sind Administratoren, außer beim Routing zwischen VLANs, nur selten daran interessiert, wie fortgeschrittene Features in einem Netzwerk genutzt werden können. Da es in diesem Buch um IP-basierte Netzwerke geht, geht es beim Switching fast immer um Ethernet-Frames und beim Routing um IP-Pakete.

Abbildung 1-1 ▶
TCP/IP-Modell

Anwendung	FTP, telnet, E-Mail, Spiele, Druck, HTTP
Transport	Transmission Control Protocol (TCP), User Datagram Protocol (UDP)
Internet	Internet Protocol (IP), ICMP, IGMP
Netzzugang	Ethernet, 802.11
Bitübertragung	Ethernet, 802.11

Switches arbeiten auf der zweiten Schicht des TCP/IP- (und OSI-)Modells und sind die Arbeitspferde der meisten Netzwerke. Die Arbeitsweise von Switches und Bridges ist im IEEE-Standard 802.1D definiert. Der Standard beschreibt auch das Verhalten anderer Schicht-2-Protokolle wie des Spanning Tree Protocol, das wir in Kapitel 3 diskutieren.

Beim Netzwerkdesign reden wir häufig von der »Zugriffsschicht« (Access Layer) oder darüber, wie Hosts mit dem Netzwerk verbunden werden. Switches und Access Points (Hubs und Kollisionsdomänen ignorieren wir hier mal) bilden den Unterbau. Neben der Weiterleitung von Ethernet-Frames basierend auf MAC-Adressen (Media Access Control) und der Verarbeitung von CRC-Prüfungen

(Cyclical Redundancy Check) bieten Switches eine Reihe sehr wichtiger Dienste an:

- Sie filtern nicht weiterzuleitenden Traffics heraus, etwa lokale Unicast-Frames.
- Sie verhindern die Weiterleitung von Kollisionen.
- Sie verhindern die Weiterleitung fehlerhafter Frames.

Switches bieten auch eine Reihe von Features, die Teil der meisten mittleren und großen Netzwerke sind:

- virtuelle lokale Netzwerke (VLANs)
- Simple Network Management Protocol (SNMP)
- Fernwartung
- Statistiken
- Port-Mirroring
- Sicherheit wie die portbasierte Authentifizierung nach 802.1X

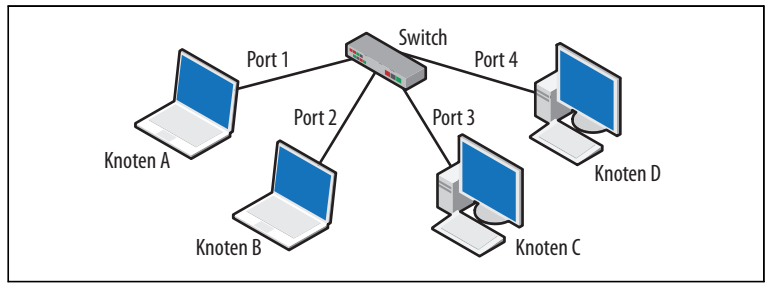
Jedes mit einem Netzwerk verbundene Gerät muss, unabhängig von seiner Spezialisierung, den Regeln dieses Netzwerks folgen. Daher befolgen Switches auch die Regeln für den Ethernet-Zugriff und die Kollisionserkennung. Sie führen auch die gleichen automatischen Verhandlungen durch, die auch die Ethernet-Hosts durchlaufen. Bei der Installation von Switches können unterschiedliche Link-Typen gewählt werden. Sie können in Punkt-zu-Punkt-Konfigurationen direkt miteinander verbunden werden oder an gemeinsam genutzte Medien oder Hosts angeschlossen werden. Je nach Lage innerhalb des Netzwerks können die Anforderungen an Performance und Sicherheit signifikant voneinander abweichen. Bei Core- oder Backbone-Switches (und Routern) kann ein extrem hoher Durchsatz erforderlich sein, während mit kritischen Elementen verbundene Switches für höhere Sicherheitsanforderungen konfiguriert sein können. Bei vielen Switches gibt es überhaupt keine Konfigurationsänderungen. Sie werden einfach ausgepackt und mit den Werkseinstellungen betrieben.

Weiterleitung basierend auf MAC-Adressen

Um Ethernet-Frames weiterzuleiten oder zu filtern, sieht sich der Switch eine Tabelle mit Quelladressen, die sogenannte SAT (Source Address Table), an, bevor er einen Frame an sein Ziel weiterleitet. Die SAT wird auch MAC-Adresstabelle oder CAM (Content Addressable Memory) genannt. Nur das in der Tabelle enthaltene Ziel empfängt die Daten. Ganz allgemein empfängt ein Switch einen Frame, liest die MAC-Adressen, führt einen Cyclical Redundancy

Check (CRC) zur Fehlerkontrolle durch und leitet den Frame schließlich an den richtigen Port weiter. Broadcast- und Multicast-Frames werden üblicherweise überall hin weitergeleitet (mit Ausnahme des ursprünglichen Quellports). Abbildung 1-2 zeigt eine typische Topologie mit einem Switch in der Mitte.

Abbildung 1-2 ▶
Einfache Switch-Topologie



Die Netzwerk-Knoten besitzen eine eindeutige MAC-Adresse, und Ethernet-Frames identifizieren die Quelle und das Ziel über diese MAC-Adressen. Eine MAC-Adresse ist ein 6-Byte-Wert wie `00:12:34:56:78:99`, der dem Host zugeordnet wird. Die SAT bildet die MAC-Adressen auf die Switch-Ports ab. Diese Tabelle hält auch die virtuellen lokalen Netzwerke (VLANs) nach, die auf dem Switch konfiguriert sind. Bei den meisten Switches liegen alle Ports standardmäßig in VLAN 1. Die SAT für das Netzwerk in Abbildung 1-2 könnte wie in Tabelle 1-1 aussehen.

Tabelle 1-1: Switch-SAT

MAC-Adresse	VLAN	Port
Knoten A MAC	1	1
Knoten B MAC	1	2
Knoten C MAC	1	3
Knoten D MAC	1	4

Ist die Adresse bekannt, wird der Frame an den richtigen Port weitergeleitet. Ist die Adresse unbekannt, wird der Frame an alle Ports (bis auf den Quellport) weitergeleitet. Das bezeichnet man als Flooding (»Fluten«). Ist die MAC-Adresse des Ziels eine Broadcast-Adresse (der Form `ff:ff:ff:ff:ff:ff`), wird der Frame ebenfalls an alle Ports (außer an den Quellport) gesendet. In vielen Fällen ist dies auch das Verhalten bei Multicast-Frames. Erinnern Sie sich daran, dass Multicast-Frames üblicherweise mit einem hexadezimalen 01 im ersten Byte beginnen. Der Bereich eines Multicast-Frames kann mithilfe des Interior Group Management Protocol (IGMP) beeinflusst werden. Switches können ein IGMP-Snooping durchführen,

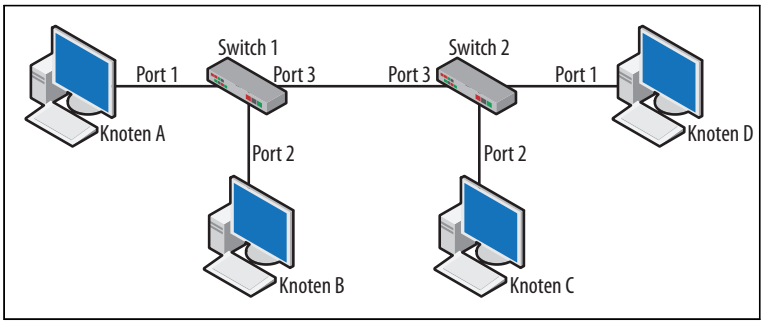
um die Ports zu ermitteln, die den Multicast-Traffic empfangen sollen. IGMP ist ebenfalls im IEEE 802.1D-Standard definiert. VLANs können die Auswirkungen des Floodings oder Broadcastings reduzieren, da man mit ihnen den Switch in kleinere logische Segmente aufteilen kann. Wir behandeln VLANs in Kapitel 4.

Abbildung 1-3 zeigt die SAT eines laufenden Cisco-Switches. Die Ausgabe wurde mit dem Befehl `show mac-address-table` des Cisco-Switches erzeugt. Der Begriff »*dynamic*« bedeutet, dass der Switch die Adressen durch die Untersuchung der von den angeschlossenen Knoten gesendeten Frames gelernt hat.

```
Non-static Address Table:
Destination Address  Address Type  VLAN  Destination Port
-------------------  ------------  ----  --------------------
0004.9b4b.5701       Dynamic          1  FastEthernet0/1
0004.9b4b.5701       Dynamic          2  FastEthernet0/1
0004.9b4b.5701       Dynamic          3  FastEthernet0/1
000e.0c76.5ad4       Dynamic          2  FastEthernet0/7
000e.0c77.20e4       Dynamic          2  FastEthernet0/1
000e.0c77.2322       Dynamic          3  FastEthernet0/1
0011.212c.15e0       Dynamic          3  FastEthernet0/23
0011.212c.15e1       Dynamic          2  FastEthernet0/13
```

◀ **Abbildung 1-3**
SAT bei Cisco-Switch

Beachten Sie, dass es hier drei VLANs gibt und dass mit Port 1 (FastEthernet0/1) verschiedene MAC-Adressen verknüpft sind. Das liegt daran, dass an diesem Punkt ein weiterer Switch angeschlossen ist. Ein Beispiel für eine solche Topologie ist in Abbildung 1-4 zu sehen. Zwei Switches sind über Port 3 an Switch 1 und Port 3 an Switch 2 miteinander verbunden. Während des normalen Datenverkehrs lernen die Switches, wo alle MAC-Ziele liegen, indem sie die Quell-MACs der Ethernet-Übertragungen festhalten.



◀ **Abbildung 1-4**
Zwei-Switch-Topologie

Bei einer solchen Topologie ist es für einen Switch unmöglich, direkt mit jedem Ziel verbunden zu sein. So besitzt Switch 2 zum

Beispiel nur die Quell-MAC aus seiner Perspektive. Aus seiner Sicht kommen alle Frames scheinbar von einem einzelnen Port (3), der mit Switch 1 verbunden ist. Umgekehrt ist es genauso. Aufbauend auf dem Wissen über SATs und dem Lernprozess, würden die SATs für die beiden Switches aussehen wie in Tabelle 1-2.

Tabelle 1-2: SAT für die Zwei-Switch-Topologie

Switch 1			Switch 2		
MAC-Adresse	**VLAN**	**Port**	**MAC-Adresse**	**VLAN**	**Port**
Knoten A	1	1	Knoten A	1	3
Knoten B	1	2	Knoten B	1	3
Knoten C	1	3	Knoten C	1	2
Knoten D	1	3	Knoten D	1	1

Sendet Knoten A Daten an Knoten D, leitet Switch 1 diese über Port 3 weiter. Switch 2 empfängt den Frame und leitet ihn an Port 1 weiter.

Abbildung 1-3 zeigt auch mehrere VLANs. Was aus den SATs und den Topologie-Diagrammen nicht hervorgeht, ist, wie Daten von einem VLAN in ein anderes wandern. Miteinander verbundene Switches, die mit VLANs konfiguriert sind, nutzen üblicherweise sogenannte *Trunk Lines*. Darüber hinaus benötigen Schicht-2-Switches einen Router oder Routingfunktionalität, um Daten zwischen VLANs weiterleiten zu können. Mit dem Aufkommen von Multiplayer-Switches verschwimmt die Grenze zwischen Routern und Switches ein wenig. VLANs und Trunks werden in Kapitel 4 ausführlich behandelt.

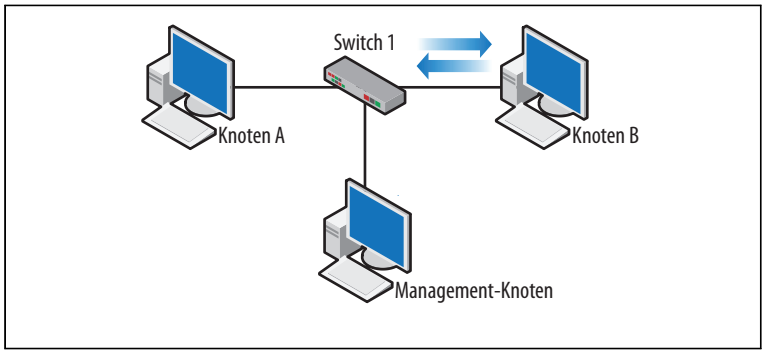
Ein anderes sehr nettes Feature von Switches ist das sogenannte Port Mirroring. Dieses Mirroring (Spiegeln) kopiert den Traffic eines Ports und sendet ihn an einen anderen. Das ist wichtig, weil in den letzten Jahren Hubs fast vollständig aus den Netzwerken verschwunden sind. Doch ohne Hub kann es schwierig sein, sich den durch das Netzwerk fließenden Datenverkehr »anzusehen«. Mittels Mirroring können Sie einen Management-Host installieren und den Traffic von jedem Port oder VLAN sammeln. Nachfolgend sehen Sie Beispiele für die Befehle, die man dazu auf einem Cisco-Switch eingibt:

```
monitor session 1 source interface Fa0/24
monitor session 1 destination interface Fa0/9 encapsulation dot1q
```

Der erste Befehl beschreibt die Quelle des zu überwachenden Datenverkehrs. Der zweite Befehl gibt nicht nur das Ziel an, sondern auch die Art der Frame-Kapselung. In diesem Fall läuft der über-

wachte Traffic über eine Trunk Line. Trunks werden in Kapitel 4 erläutert. Mirroring-Befehle können auch die Richtung des gewünschten Datenverkehrs angeben. Es ist also möglich, den Datenverkehr auszuwählen, der von oder zu einem bestimmten Host läuft. Typischerweise sind beide Richtungen voreingestellt.

Abbildung 1-5 zeigt ein Beispiel, bei dem die Knoten A und B miteinander kommunizieren, während der Netzwerk-Admin wissen will, was da passiert. Der Traffic von Knoten B wird also auf den Management-Knoten gespiegelt. Da die Kommunikation zwischen den Knoten A und B erfolgt, reicht ein Port aus, der mit einem von beiden verbunden ist.



◀ **Abbildung 1-5**
Port-Mirroring

Routing: Wege finden

Beim Aufbau von Netzwerken teilen wir das Routing üblicherweise in zwei Komponenten auf: Host und Router. Router verarbeiten Traffic, der zwischen Netzwerken fließt, doch Hosts treffen viele Entscheidungen, lange bevor die Pakete das Netzwerk überhaupt erreichen. Die meisten Routingprotokolle, die genutzt werden, um Wege zu den Zielen zu finden, sind allerdings routerbasiert.

Hosts werden typischerweise auf eine von zwei Arten konfiguriert: statisch mit IP-Adresse, Standard-Gateway und Domain-Name-Server, oder sie werden mit Werten gefüttert, die sie über das Dynamic Host Configuration Protocol (DHCP) gelernt haben. Hosts senden den gesamten Traffic, der nicht für das lokale Netzwerk bestimmt ist, an das Standard-Gateway. Sie setzen darauf, dass das Gateway die Pakete an das Ziel weiterleiten kann. Eine meiner Lieblingsfragen lautet: »Was macht ein Host als Erstes, bevor er ein Paket sendet?« Vor allem anderen muss ein Host zuerst seine Routingtabelle verarbeiten. Kapitel 2 widmet sich dem hostbasierten Rou-

ting. Historisch betrachtet gab es einige Netzwerktechniken, bei denen die Hosts aktiver waren. Zum Beispiel nutzte IBMs Token Ring sogenannte Discovery-Frames, um Zielknoten in anderen Netzwerksegmenten oder Ringen zu finden. Allerdings ist das primär eine Schicht-2-Funktion und nicht Teil moderner Ethernet- und IP-basierter Netzwerke. In den letzten Jahren kann man im Bereich Ad-hoc-Netzwerke eine Rückkehr zur Nutzung von Hosts für Routingfunktionen beobachten.

Ad-hoc-Routing läuft üblicherweise nicht in traditionellen Netzwerk-Infrastrukturen. Anwendungsfälle sind Sensor-Netzwerke, die Kommunikation im Gefecht und Katastrophenszenarien, bei denen die Infrastruktur zerstört wurde. In diesen Fällen übernehmen die Knoten die Weiterleitung von Daten an andere Knoten. Verwandte Ideen sind Ad-hoc-Anwendungen und 802.11-Ad-hoc-Netzwerke. Wichtig ist dabei, dass im 802.11-Standard Knoten die Verbindung zu einem Ad-hoc-Netzwerk herstellen, aber keinen Traffic an andere Knoten weiterleiten können. Liegt ein drahtloser Knoten nicht im Bereich des Quell-Hosts, schlägt die Übertragung fehl.

Ad-hoc-Routingprotokolle wurden entworfen, um eben dieses Problem zu lösen, indem sie es den Hosts ermöglichen, die Routing/Forwarding-Funktionen zu übernehmen. Interessante Probleme treten auf, wenn der »Router« nicht mit dem Netzwerk verbunden ist: Dinge wie die Verschiebung von Drahtlos-Knoten, Stromsparfunktionen, Zugriffsfähigkeit und Speicher können betroffen sein. Darüber hinaus ist die Anwendung von Bedeutung. Handelt es sich bei den Knoten um Sensoren, die nur sehr wenige Ressourcen besitzen? Bewegen sie sich schnell? Diese Anforderungen haben zur Entwicklung von Ad-hoc-Routingprotokollen wie Ad hoc On Demand Distance Vector (AODV), Fisheye State Routing (FSR) und Optimized Link State Routing (OLSR) geführt.

Doch diese Ideen gehen über den Rahmen dieses Buches hinaus. Der eigentliche Punkt ist, dass Hosts und die Routingtabelle des Hosts bei der Verarbeitung von Paketen sehr aktiv sind. Historisch gesehen waren die Knoten in einigen Netzwerken viel stärker involviert, und wenn man die Ad-hoc-Routingprotokolle als Zeichen werten will, dann sind diese Zeiten noch lange nicht vorbei.

Routinggeräte

Router arbeiten in der Netzwerkschicht des TCP/IP-Modells und verarbeiten IP-Adressen basierend auf ihrer Routingtabelle. Die Hauptaufgabe eines Routers besteht darin, Daten basierend auf der Zieladresse in einem IP-Paket an die Zielnetzwerke weiterzuleiten.

Router können außerdem MAC-Adressen (insbesondere die eigenen) mithilfe des Address Resolution Protocol (ARP) auflösen. Es ist wichtig, daran zu denken, dass die Schicht-2-Frames und MAC-Adressen nicht über den Router hinaus existieren. Das bedeutet, dass ein Ethernet-Frame zerstört wird, sobald er den Router erreicht. Wird er in einem Netzwerk betrieben, kann der Router als Standard-Gateway für die Hosts dienen, wie das in den meisten Heimnetzwerken der Fall ist. Ein Router kann als Verbindung (Hop) zu anderen Routern installiert werden, ohne direkt mit irgendwelchen Hosts verbunden zu sein. Neben dem Routing können Router eine Reihe weiterer Aufgaben übernehmen, etwa die Übersetzung von Netzwerkadressen (Network Address Translation), die Verwaltung von Zugriffskontrolllisten, die Terminierung virtueller privater Netzwerke oder QoS (Quality of Service).

Die grundlegende Router-Funktionalität setzt sich aus drei Hauptbestandteilen zusammen:

- Routingprozess
- Routingprotokolle
- Routingtabelle

Beim Routingprozess geht es darum, wie IP-Pakete von einem Port zu einem anderen bewegt werden. Die Routingtabelle enthält die Informationen, die vom Routingprozess genutzt werden. Routingprotokolle wie das Routing Information Protocol (RIP) oder Open Shortest Path First (OSPF) werden zur Kommunikation mit anderen Routern verwendet. Diese Kommunikation kann in der »Installation« von Routen in der Routingtabelle münden, die dann vom Routingprozess genutzt werden. Bei der Konfiguration eines Routers wird die Routingtabelle aufgebaut, indem man Interfaces aktiviert und diese mit IP-Adressen versorgt. Eine einfache Cisco-Routingtabelle ist in Abbildung 1-6 zu sehen.

```
Gateway of last resort is not set

C    192.168.15.0/24 is directly connected, FastEthernet0/1
C    192.168.20.0/24 is directly connected, FastEthernet0/0
```

◀ Abbildung 1-6
Routingtabelle eines Routers

Bei der Verarbeitung von Paketen gehen Router die Routingtabelle durch und suchen den bestmöglichen Weg. Die Routingtabelle in Abbildung 1-6 zeigt, dass der Router zwei Netzwerke kennt: `192.168.15.0` und `192.168.20.0`. Beachten Sie, dass dieser Router kein Standard-Gateway, also keinen »letzten Ausweg« kennt. Das bedeutet, dass der Router nicht weiß, wie er die Pakete zustellen

soll, wenn die Zieladresse nicht in diesen beiden Netzwerken liegt. Wenn Sie sich selbst »Ah, ICMP destination unreachable« sagen hören, können Sie sich einen Orden anheften.

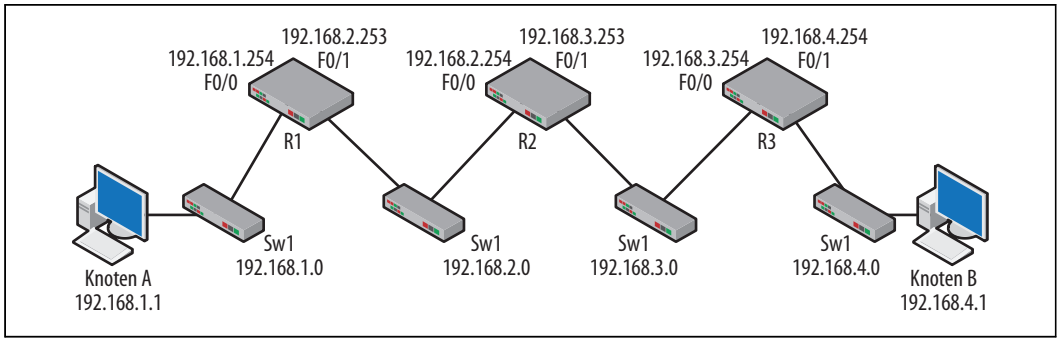
Routingtabellen können aus unterschiedlichen Arten von Routen bestehen: aus Direktverbindungen, statischen und dynamischen Routern. Zwei Direktverbindungen sehen Sie in Abbildung 1-6. Das sind die Netzwerke, in denen der Router ein Interface besitzt. Sie enthalten den Buchstaben »C« sowie die jeweilige Schnittstelle, etwa `FastEthernet0/1`. Direkt verbundene Routen haben Vorrang vor allen anderen Routen.

Tipp Das 0/1 des Interfaces ist ein Bezeichner für das Blade und den Port im Router-Chassis.

Statische Routen

Statische Einträge werden auf dem Router durch den Netzwerkadministrator von Hand vorgenommen. Für spezielle Ziele und bei sehr kleinen oder stabilen Netzwerkumgebungen können manuell konfigurierte statische Routen sehr erfolgreich verwendet werden. Bei der Verwendung statischer Routen hat der *Netzwerkadministrator* den Weg festgelegt, der zu einem bestimmten Zielnetzwerk verwendet werden soll. Die statische Route ersetzt aufgrund der administrativen Distanz (die wir später noch diskutieren werden) jeden durch das Routingprotokoll gelernten Weg.

Eine weitere, wichtige und zentrale Routingidee ist der *nächste Hop*. Der nächste Hop ist ein Router, der aus Sicht eines bestimmten Routers einen Schritt näher am Ziel liegt. Der nächste Hop ist der Router, an den die Pakete als Nächstes gesendet werden. In vielen Netzwerken wird eine Reihe von nächsten Hops verwendet. Eine kleine Routingtopologie ist in Abbildung 1-7 zu sehen. Aus der Perspektive von R1 ist also R2 der nächste Hop, um die Netzwerke `192.168.3.0` und `192.168.4.0` zu erreichen.



▲ **Abbildung 1-7**
Kleine Routingtopologie

Diese Topologie besteht aus drei Routern, die wie abgebildet über die Switches miteinander verbunden sind. Es gibt verschiedene Möglichkeiten, eine solche Topologie zu emulieren, doch diese Konfiguration wurde der Klarheit halber gewählt. Am Anfang ist noch nichts konfiguriert, nur die Interfaces sind »oben« und haben IP-Adressen erhalten. Um ein Interface hochzubringen, muss es den Befehl `no shutdown` erhalten und einen Link-Impuls haben. Die Routingtabellen auf den Routern enthalten nur die direkt verbundenen Routen. Jeder Router ist sich nur der beiden Netzwerke bewusst, für die er Interfaces hat. Tabelle 1-3 zeigt die Routingtabellen zu diesem Zeitpunkt.

Tabelle 1-3: Ausgangs-Routingtabellen

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1

Aus dieser Tabelle wird deutlich, dass die Router kein vollständiges Bild des gesamten Netzwerks haben. Beispielsweise ist Knoten A mit Switch 1 verbunden und versucht, Knoten B an Switch 4 zu erreichen. Nach der Verarbeitung seiner Host-Routingtabelle (siehe Kapitel 2) leitet er den Traffic an sein Standard-Gateway (`192.168.1.254`) an R1 weiter. R1 betrachtet nun seine Routingtabelle und entdeckt, dass er nur Einträge für die Netzwerke der linken Seite der Topologie besitzt. Ohne Wissen um das Zielnetzwerk gibt R1 eine »ICMP destination unreachable«-Meldung aus.

Tipp

Die Netzwerke `192.168.1.0` und `192.168.4.0` bezeichnet man als Stub-(Stummel-)Netzwerke, weil es nur einen Weg hinein oder hinaus gibt.

Wie wird dieses Problem gelöst? Bei kleinen Netzwerken wie diesen kann der Netzwerkadministrator Routingbefehle an die Router ab-

setzen, um sie mit zusätzlichen Forwarding-Informationen zu versorgen. Das wären die statischen Routen. Bei Cisco-Routern wird der Befehl `ip route` verwendet. Er besitzt drei Felder, die vom Netzwerkadministrator aufgefüllt werden müssen:

```
ip route zielnetzwerk ziel-netzwerkmaske nächster-hop-ip-adres-
se(forwarding-router-interface)
```

Beispielsweise könnten Sie R1 mit den folgenden Befehlen mitteilen, wie er die Netzwerke `192.168.3.0` und `192.168.4.0` erreicht:

```
ip route 192.168.3.0 255.255.255.0 192.168.2.254
ip route 192.168.4.0 255.255.255.0 192.168.2.254
```

Die Befehle sind bis auf das Zielnetzwerk identisch. Hier einige wichtige Punkte: Das letzte Feld, das das weiterleitende Router-Interface (`192.168.2.254`) festlegt, ist ein benachbarter Router, der von R1 erreicht werden kann. Mit diesen beiden Befehlen legen Sie fest, dass der Traffic für diese beiden Netzwerke an R2 gesendet werden soll. Die Maske ist ebenfalls die Maske des Zielnetzwerks und nicht die lokal verwendete Maske. Diese Masken können unterschiedlich sein. Die korrekte Form wird rekursive Route genannt.

Nachdem die Befehle auf R1 eingegeben wurden, sieht die aktualisierte Routingtabelle so aus wie in Tabelle 1-4:

Tabelle 1-4: Aktualisierte R1-Routingtabelle

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
S 192.168.3.0 via 192.168.2.254		
S 192.168.4.0 via 192.168.2.254		

Das ist zwar eine Verbesserung, löst aber nur einen Teil des Problems. R1 weiß nun, dass für die Netzwerke gedachter Traffic an R2 gehen muss, doch was macht R2 als Nächstes? Im Falle des `192.168.3.0`-Netzwerks ist alles gut, da es direkt mit R2 verbunden ist. R2 kann ARP für die Hosts nutzen, da diese im gleichen Netzwerk liegen. Ist der Traffic jedoch für `192.168.4.0` gedacht, braucht R2 etwas Hilfe vom Administrator in Form des folgenden Befehls:

```
ip route 192.168.4.0 255.255.255.0 192.168.3.254
```

Die Routingtabelle wird entsprechend aktualisiert, und wir können erleichtert aufatmen, weil es die Pakete schließlich auch in das `192.168.4.0`-Netzwerk geschafft haben.

Tabelle 1-5: Aktualisierte R2-Routingtabelle

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
S 192.168.3.0 via 192.168.2.254	S 192.168.4.0 via 192.168.3.254	
S 192.168.4.0 via 192.168.2.254		

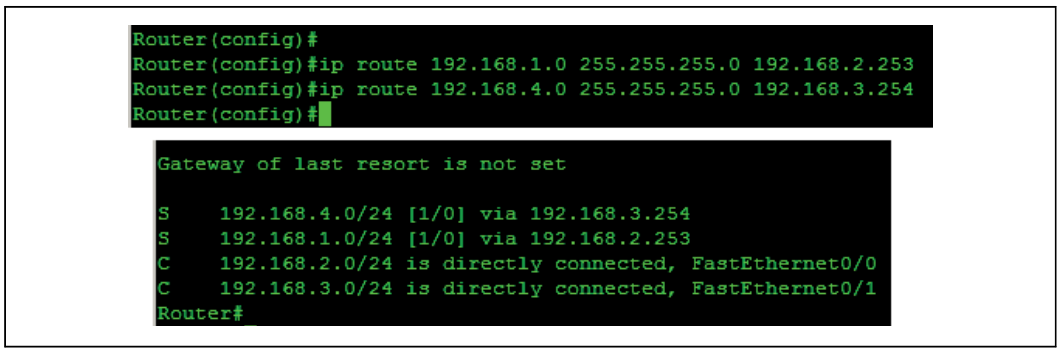
Das Zielnetzwerk zu erreichen ist nur die halbe Miete – Datenpakete müssen es auch zurück schaffen. Wenn wir uns die Routingtabelle auf R3 ansehen, erkennen wir, dass der Router nicht weiß, wo er das `192.168.1.0`-Netzwerk finden kann. Das Paket von Knoten A gelangt zwar dort hin, doch wenn Knoten B antworten will, bekommt er von R3 nur eine »ICMP destination unreachable«-Meldung. Für Knoten A sieht es so aus, als wäre die Übertragung nie beantwortet worden. Um die Sache abzuschließen, müssen `ip route`-Befehle für alle unbekannten Netzwerke auf den jeweiligen Routern abgesetzt und die Routingtabellen aktualisiert werden. Sobald alle `ip route`-Befehle eingegeben wurden, sieht die Routingtabelle aus wie in Tabelle 1-6.

Tabelle 1-6: Vollständige Routingtabellen

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
S 192.168.3.0 via 192.168.2.254	S 192.168.1.0 via 192.168.2.253	S 192.168.1.0 via 192.168.3.253
S 192.168.4.0 via 192.168.2.254	S 192.168.4.0 via 192.168.3.254	S 192.168.2.0 via 192.168.3.253

Die Routingtabelle für R2 und die auf R2 eingegebenen `ip route`-Befehle sind in Abbildung 1-8 zu sehen.

▼ Abbildung 1-8
R2-Routingtabelle samt Befehlen für statische Routen



Bei diesen Routingtabellen konnten alle Zielnetzwerke erreicht werden, weil sie entweder direkt verbunden waren oder weil es eine statische Route zu einem Nachbar-Router gibt, der eventuell weiterhelfen kann. Ich sage »eventuell«, weil wir bei der Verwendung statischer Routen annehmen, dass der andere Router etwas über den Weg zum Ziel weiß. Das ist nicht immer der Fall, was Sie erkennen konnten, als die Routingtabellen noch nicht vollständig gefüllt waren.

Tipp Es gibt verschiedene Optionen im Bezug auf die Argumente des `ip route`-Befehls, und es gibt Zeiten, in denen die hier gezeigte Verwendung modifiziert werden muss. Serielle Links sind ein gutes Beispiel. Bei ihnen muss das letzte Feld ein Interface sein und nicht die IP-Adresse des nächsten Hops.

Typische Fehler

Schaut man sich die in Abbildung 1-8 gemachten Änderungen genauer an, gibt es zwei typische Fehler bei der Konfiguration statischer Routen. Wir wollen sie uns aus der Perspektive von R2 ansehen. Hier ist ein typischer Fehler:

```
ip route 192.168.1.0 255.255.255.0 192.168.2.254
```

Dieser Befehl fordert den Router auf, Traffic an sich selbst weiterzuleiten. Letztendlich besagt er: »R2 weiß nicht, wo das `192.168.1.0`-Netzwerk ist, also senden wir die Daten an R2.« Das erscheint dem Router nicht sinnvoll, und er antwortet üblicherweise mit einer Meldung wie in Abbildung 1-9. Der Netzwerkadministrator und der Router glotzen sich kurz an, und dann macht der Administrator üblicherweise den zweiten typischen Fehler. Dieser tritt auch ein, wenn Adressen falsch eingegeben werden. Die korrekte Form ist in Abbildung 1-8 zu sehen.

```
Router(config)#
Router(config)#ip route 192.168.1.0 255.255.255.0 192.168.2.254
%Invalid next hop address (it's this router)
Router(config)#
```

Abbildung 1-9 ▲ Fehlermeldung bei zirkulärem Routing

Beim zweiten Fehler wird keine IP-Adresse für den Router eingetragen, sondern nur eine physische Schnittstelle. Das führt zu einer höheren Last auf dem Router und ist üblicherweise internen Routingprotokollen vorbehalten. Der Befehl und die daraus resultierende Routingtabelle sind in Abbildung 1-10 zu sehen. Obwohl es statische Routen gibt, deutet die Routingtabelle an, dass die Netz-

werke 192.168.1.0 und 192.168.4.0 direkt miteinander verbunden sind. Die Topologie zeigt, dass das offensichtlich nicht der Fall ist.

```
ip classless
ip route 192.168.1.0 255.255.255.0 FastEthernet0/0
ip route 192.168.4.0 255.255.255.0 FastEthernet0/1
ip http server
```

```
Gateway of last resort is not set

S    192.168.4.0/24 is directly connected, FastEthernet0/1
S    192.168.1.0/24 is directly connected, FastEthernet0/0
C    192.168.2.0/24 is directly connected, FastEthernet0/0
C    192.168.3.0/24 is directly connected, FastEthernet0/1
Router#
```

◀ **Abbildung 1-10**
Fehler 2

Der Grund für die höhere Last ist, dass der Befehl nicht spezifisch genug ist und der Router tatsächlich keine Ahnung hat, wo er den Traffic hinschicken soll. Das ist so, als würde jemand einen Brief schreiben, ihn mit einer Adresse versehen und ihn dann einfach aus der Tür werfen, in der Hoffnung, dass er sein Ziel schon erreichen wird. Wirklich interessant ist die Auswirkung auf den Netzwerk-Traffic. Der ARP- (Address Resolution Protocol-)Traffic ist auf das lokale Netz- oder Subnetzwerk beschränkt. Das bedeutet, dass ARP-Meldungen von Routern generell nicht weitergeleitet werden und dass Hosts kein ARP für Knoten nutzen, die nicht in ihrem Netzwerk liegen. Eine Ausnahme findet man bei Proxy-ARP, sie wird aber nur selten genutzt. Schließlich haben MAC-Adressen über das eigene Netzwerk hinaus keine Bedeutung. Doch sehen Sie sich an, was passiert, wenn die Befehle aus Abbildung 1-10 verwendet werden. Abbildung 1-11 zeigt, dass R3 (192.168.3.254) einen ARP-Request für 192.168.1.1 sendet, einen Knoten in einem anderen Netzwerk. Das widerspricht allen grundlegenden Verhaltensweisen und ist schlicht falsch. Mir wird ganz unbehaglich, wenn ich das sehe.

▼ **Abbildung 1-11**
Nicht-lokaler ARP-Traffic

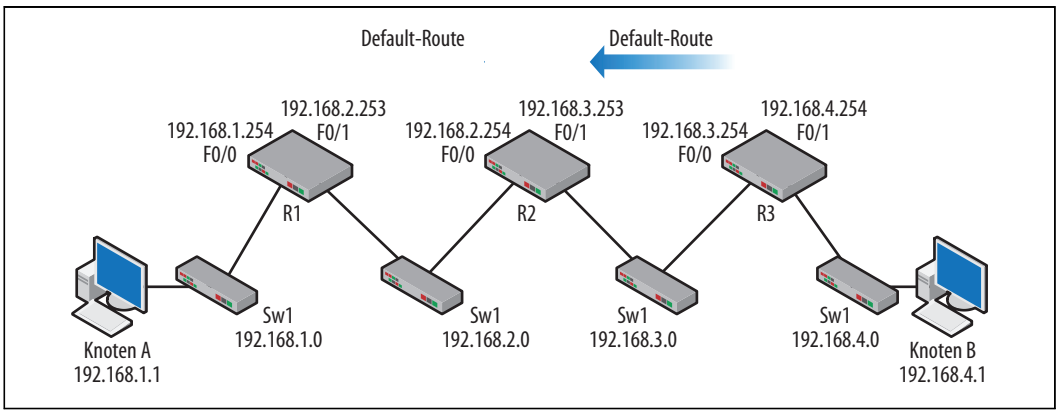
```
⊞ Ethernet II, Src: Cisco_2c:0c:80 (00:11:21:2c:0c:80), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
⊟ Address Resolution Protocol (request)
    Hardware type: Ethernet (0x0001)
    Protocol type: IP (0x0800)
    Hardware size: 6
    Protocol size: 4
    Opcode: request (0x0001)
    [Is gratuitous: False]
    Sender MAC address: Cisco_2c:0c:80 (00:11:21:2c:0c:80)
    Sender IP address: 192.168.3.254 (192.168.3.254)
    Target MAC address: 00:00:00_00:00:00 (00:00:00:00:00:00)
    Target IP address: 192.168.1.1 (192.168.1.1)
```

Standard-(Default-)Routen

Es kommt oft vor, dass mehrere Ziele über den gleichen Weg erreicht werden können. In solchen Fällen kann die Routingtabelle anwachsen, obwohl viele Routen einige Felder gemeinsam haben. Das gilt für die Routingtabelle von R1 und R3. Einträge in der Routingtabelle, die den gleichen Weg nutzen, können durch eine kleinere Gruppe von Routen ersetzt werden. Die besten Beispiele sind Standard-(Default-)Routen und die Aggregation. Aggregation oder Routen-Zusammenfassung (Route Summarization) ist eine Technik, die die Anzahl der Einträge in einer Routingtabelle reduziert, indem sie die Präfixlänge kürzt. Auf diese Weise werden eine Reihe von Zielen in einem einzelnen Eintrag zusammengefasst.

Die Default-Route ist ein Sonderfall einer statischen Route. Normalerweise denken wir an Standard-Gateways oder Router für Hosts. Router können ebenfalls Standard-Gateways besitzen. Ist die Routingtabelle (wie bei einem Host) ausgeschöpft und wurde kein Treffer für das Ziel gefunden, dann wird die Default-Route verwendet. Im Cisco-Sprachgebrauch ist das das »Gateway zum letzten Ausweg« (»gateway of last resort«). Genau wie bei statischen Routen nimmt der Netzwerkadministrator an, dass der Router am nächsten Hop etwas weiß, das der aktuelle Router nicht weiß: wie man entweder zum Ziel oder zum nächsten Hop kommt. Abbildung 1-12 zeigt die Topologie mit den Default-Route-Kandidaten, basierend auf den Informationen aus Tabelle 1-6.

Abbildung 1-12 ▼
Default-Routen



Für R1 sind alle nicht direkt angeschlossenen Ziele nur erreichbar, wenn man den Traffic an `192.168.2.254` weiterleitet. Für R3 sind alle nicht direkt verbundenen Ziele nur erreichbar, wenn der Traffic an `192.168.3.253` weitergeleitet wird. Daher können einige Tabelleneinträge durch eine Default-Route ersetzt werden. Für einen Router

wird die Default-Route (oder das »Gateway des letzten Auswegs«) mit einer Reihe spezieller Argumente im `ip route`-Befehl angegeben. Anstatt ein Zielnetzwerk und eine Zielnetzwerkmaske anzugeben, verwenden Default-Routen nur Nullen. Erinnern Sie sich daran, dass bei der Verarbeitung von Routingtabellen mit Masken die UND-Verknüpfung einer IP-Adresse mit der Maske `0.0.0.0` zu `0.0.0.0` führt. Das bedeutet, dass jedes Ziel nur Nullen ergibt (`0.0.0.0`), ebenso wie die UND-Verknüpfung für diese `ip route`-Zeile nur Nullen ergibt, wodurch sie jedes Ziel erkennt:

```
ip route 0.0.0.0 0.0.0.0 forwarding router interface
```

Für R1:

```
ip route 0.0.0.0 0.0.0.0 192.168.2.254
```

Und für R4:

```
ip route 0.0.0.0 0.0.0.0 192.168.3.253
```

Die Routingtabellen sehen dann aus wie in Tabelle 1-7.

Tabelle 1-7: Um Default-Routen aktualisierte Routingtabellen

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
S 0.0.0.0/0 via 192.168.2.254	S 192.168.1.0 via 192.168.2.253	S 0.0.0.0/0 via 192.168.3.253
	S 192.168.4.0 via 192.168.3.254	

Wieder gibt es einige wichtige Dinge zu bemerken. Während die Routingtabellen für R1 und R3 verbessert wurden, besitzt R2 noch die gleiche Anzahl von Routen. In diesem Fall würde eine Default-Route für R1 oder R3 nicht helfen, weil R2 immer noch eine weitere Route für das Netzwerk in der anderen Richtung benötigt. Darüber hinaus würden wir eine Routingschleife riskieren. Letztlich erscheint einem die Reduzierung von vier auf drei Routen in R1 und R3 nicht als besondere Verbesserung, doch dies ist ein kleines Netzwerk. Produktionsnetzwerke sind wesentlich größer und haben Hunderte von Routen.

Dynamische Routen

Dynamische Routen werden über Routingprotokolle wie das Routing Information Protocol (RIP) oder Open Shortest Path First (OSPF) erlernt. Beim Aufbau eines Netzwerks ist der Ansatz, mit dem man das Routing handhabt, eine wichtige Entscheidung. Statische Routen verringern die Last, doch Änderungen in der Topologie können nicht schnell verarbeitet werden. Ändert sich der Weg zu

einem Ziel oder ist ein Router offline, gehen Wege oder Routen verloren. Statische Routen sind vor Fehlern des Admins in keiner Weise geschützt. Üblicherweise werden statische Routen verwendet, wenn die Topologie stabil und die Netzwerkarchitektur recht einfach ist. Mit anderen Worten: wenn die Bedingungen im Netzwerk gut verstanden werden. Wir gehen häufig davon aus, dass eine Route korrekt sein muss, wenn der Admin sie installiert. Dynamische Routingprotokolle können uns vor solchen Topologie-Änderungen schützen und Fehler an der Tastatur vermeiden. Die meisten Routingprotokolle bieten auch Schutz vor Routingschleifen und alten, fehlerhaften Informationen. Viele können auch die Last verteilen (Load Balancing) und mehrere Wege zum Ziel finden.

Routingprotokolle

Bevor wir uns in den späteren Kapiteln die einzelnen Routingprotokolle ansehen, müssen wir die Arten bzw. Charakteristika von Protokollen diskutieren. Die Idee ist, das für die Aufgabe geeignete Protokoll zu wählen, und zu diesem Zweck müssen wir uns den Algorithmus und den Betrieb detailliert ansehen. Protokolle kann man auf verschiedene Weise betrachten bzw. definieren.

Single- versus Multipath

Routingprotokolle verwenden einen Algorithmus, um den besten Weg (Pfad) zum Ziel zu ermitteln. Gibt es nur einen Weg, ist die Entscheidung einfach. Gibt es mehrere Wege zum Ziel, hat das Routingprotokoll die Wahl: Es kann ausschließlich den bestmöglichen Weg verwenden und andere Wege ignorieren, bis diese gebraucht werden, oder es kann mehrere Wege (Pathways) zum Ziel installieren. Die erste Variante wird *Singlepath*-Protokoll genannt. Es kann auch sein, dass zwei Pfade in allen Belangen gleich sind und der Router keine Entscheidung treffen kann, welche Wahl die bessere ist. Das Protokoll kann die Daten dann über beide Pfade senden. In diesem Fall könnte das Protokoll eine Art Lastverteilung (Load Balancing) vornehmen, um den Netzwerkdurchsatz zu verbessern. Man spricht dann von von einem *Multipath*-Protokoll. Schließlich muss man sich noch mit Backup-Pfaden und (falls ein Weg nicht verfügbar ist) mit den Ausfallsicherungsfähigkeiten des Protokolls beschäftigen.

Intern versus extern

Die meisten Protokolle unterliegen bestimmten Beschränkungen. Ein Beispiel ist das Routing Information Protocol (RIP), das Netz-

werke mit mehr als 15 Hops nicht verarbeiten kann. Die Protokolle sind auch so entworfen, dass sie in ihre Berechnungen bestimmte Netzwerkparameter wie Kosten oder Auslastung mit einbeziehen. Es kann daher durchaus sein, dass ein bestimmtes Protokoll für eine Netzwerk-Topologie völlig ungeeignet ist. Diejenigen, die für eine Gruppe von Netzwerken unter einer einzelnen administrativen Kontrolle (ein autonomes System) entworfen wurden, nennt man *interne* Routingprotokolle. Wir werden in späteren Kapiteln sehen, dass einige interne Routingprokolle nur für kleine Netzwerkgruppen geeignet sind. Diejenigen, die für wesentlich größere Topologien wie WAN-Konnektivität entworfen wurden, oder solche, die von ISPs genutzt werden, bezeichnet man als *externe* Protokolle. Externe Protokolle verbinden autonome Systeme miteinander. Das Border Gateway Protocol (BGP) ist ein solches externes Routingprotokoll.

Flach versus hierarchisch

Bei der Implementierung eines Routingprotokolls haben Router bestimmte Aufgaben zu erledigen, etwa Routinginformationen anzubieten, Topologieänderungen zu verarbeiten und den besten Pfad zu ermitteln. Führen alle Router die gleichen Aufgaben durch, spricht man von einem *flachen* Protokoll. Das ist bei RIP der Fall. Werden einem Teil der Router hingegen andere Funktionen zugewiesen, spricht man von einer *hierarchischen* Betriebsform. Zum Beispiel definieren einige Protokolle Backbone- und Nicht-Backbone-Bereiche des Netzwerks. Der Traffic fließt häufig von den Nicht-Backbone- in die Backbone-Bereiche. Protokolle ziehen häufig Grenzen um diese Bereiche, die als Domänen (oder Areas) bezeichnet werden. Gleichrangige (Peer-)Router kommunizieren innerhalb der Domäne, und Backbone-Router kommunizieren zwischen den Domänen. OSPF wird aufgrund seiner bereichsbasierten Organisation als hierarchisch betrachtet. Alle OSPF-Router beherrschen das Forwarding innerhalb eines Bereichs. Einige Router beherrschen das Forwarding zwischen einzelnen Bereichen und verfügen über zusätzliches Wissen zur Gesamt-Topologie.

Link-State versus Distanzvektor

Diese beiden Begriffe bezeichnen den Algorithmus, der verwendet wird, um die zu nutzenden Routen zu ermitteln. Distanzvektor-Protokolle werden auch (nach ihren Erfindern) Bellman-Ford-Protokolle genannt. Vielleicht erinnern Sie sich noch aus dem Physikunterricht daran, dass ein Vektor ein Objekt ist, das einen Betrag und eine Richtung beschreibt. Ein Beispiel ist ein Läufer, der sich

mit 10 km/h Richtung Norden bewegt. *Distanzvektor*-Routingprotokolle verwenden die gleiche Idee, d.h., sie beschreiben die Distanz zum Ziel, üblicherweise in Hops (Anzahl der Router), und eine Richtung in Form der nächsten Hop-IP-Adresse oder des zu verwendenden Interfaces. Das Zielnetzwerk ist also X Hops weit weg, und die Pakete werden an einen bestimmten Router gesendet. Benachbarte Router senden sich einen Teil ihrer Routingtabelle gegenseitig zu und senden sich dann regelmäßig Updates. Viel mehr Informationen als Hop-Zahl und Richtung gibt es allerdings nicht. Es ist daher schwierig, eine Entscheidung anhand der Qualität des Pfades zu treffen. RIP ist ein Distanzvektor-Protokoll. Distanzvektor-Protokolle sind (im Vergleich zu Link-State-Protokollen) generell langsam, wenn es darum geht, sich »der Topologie anzugleichen«, d.h., einen stabilen Zustand der Topologie herzustellen, nachdem sich etwas geändert hat.

Link-State-Protokolle nutzen sehr viel mehr Details über die Links (oder Verbindungen) zwischen den Routern, um qualifiziertere Entscheidungen zu treffen. Wenn beispielsweise zwei Pfade den gleichen Distanzvektor zum Ziel aufweisen, der eine Pfad aber auf 1-Gbps-Ethernet setzt, während der andere ein langsameres Frame Relay nutzt, dann wird der erste Pfad gewählt, selbst wenn die Anzahl der Hops gleich ist. Diese Routinginformation wird ebenfalls durch die gesamte Topologie verteilt, um die Konvergenz zu beschleunigen. Nachdem diese Information verteilt wurde, halten die Router regelmäßig Kontakt über »Hello«-Meldungen, die andeuten, dass nichts weiter passiert ist. Aus diesen Gründen erfolgt die Konvergenz bei Link-State-Protokollen schneller. Die Protokolle basieren auf Dijkstras Algorithmus zur Bestimmung des kürzesten Pfades zwischen den Punkten eines Graphen. OSPF ist ein Beispiel für ein Link-State-Routingprotokoll.

Ein Protokoll wie RIP lässt sich also wie folgt charakterisieren: dynamisch, routerbasiert, Singe-Path, intern, flach und Distanzvektor. Warum RIP diese Charakteristika aufweist, behandeln wir in Kapitel 5. OSPF ist dann dynamisch, routerbasiert, Multipath, intern, hierarchisch und Link-State. Einen detaillierten Blick auf OSPF werden wir in Kapitel 6 werfen.

Eine Route wählen oder installieren

Da eine Routingtabelle über dynamisch gelernte Routen aufgebaut wird, muss der Router entscheiden, ob er eine Route in der Tabelle installiert. Bei statischen Routen hat der Router keine Wahl. Während Pakete vom Router empfangen werden, muss er außerdem

entscheiden, welche Route für ein bestimmtes Ziel die beste ist. Bei beiden Entscheidungen werden drei Werte (in der Reihenfolge ihrer Bedeutung) miteinander verglichen: Präfixlänge, administrative Distanz und die Metrik. Diese drei werden üblicherweise im Kontext von Cisco-Routern diskutiert, doch andere Hersteller verwenden ähnliche Prozesse und Werte beim Aufbau der Routingtabelle und bei der Entscheidungsfindung.

Präfixlänge

Die *Präfixlänge* basiert auf der Anzahl der Bits in der Maske, da die Maske die Netzwerkadresse bestimmt. Je größer die Zahl der Einsen in der Maske ist, desto größer ist die Präfixlänge. So hat beispielsweise die IP-Adresse `192.168.1.5` mit der Maske `255.255.255.0` die Netzwerkadresse `192.168.1.0`. Die Präfixlänge ist hier also 24. Die gleiche IP-Adresse mit der Maske `255.255.0.0` hat eine Präfixlänge von 16, und die Netzwerkadresse lautet `192.168.0.0`. Beim Aufbau einer Routingtabelle und beim Forwarding von Paketen werden längere Präfixe bevorzugt, weil sie ein Paket näher an das Ziel bringen. Wenn Sie zum Beispiel jemandem im Osten der USA einen Brief schicken wollen, aber nur wissen, dass er in Boston wohnt, dann würde das Postflugzeug den Brief einfach über der Stadt abwerfen und Sie hoffen, dass er sein Ziel erreicht. Schreibt man den Straßennamen auf den Brief, kommt er seinem Ziel ein Stück näher. Und fügen Sie noch die Hausnummer hinzu, erreicht er schließlich sein Ziel. Die Adresse wird also länger und länger.

Ähnliches geschieht, wenn Sie mir hier am RIT (keine DoS-Angriffe bitte) ein Paket schicken wollen. Routingtabellen-Einträge mit der Netzwerkadresse `129.21.0.0` bringen es in diesen Bereich, doch das RIT ist groß. Router ermitteln letztlich das richtige Subnetz anhand eines längeren Präfix und bringen das Paket näher ans Ziel. Die Präfixlänge ist bei diesem Prozess der wichtigste Punkt.

Administrative Distanz

Die zweite Erwägung ist die *administrative Distanz*. Manchmal erhalten Router Informationen von verschiedenen Protokollen. Wie ermittelt der Router, welche Informationen die besten sind, wenn die Präfixlänge gleich ist? Sie haben von verschiedenen Freunden etwas über zwei neue Restaurants gehört. Aus Erfahrung wissen Sie, welcher Ihrer Freunde in Bezug auf Essen die besseren Tipps gibt. Auch bei Routingprotokollen sind einige besser als andere. Die administrative Distanz ist eine (Kenn-)Zahl, die den Wert der Information beschreibt, der über ein Routingprotokoll erlernt oder bereits in der Routingtabelle vorhanden ist.

Jedes Routingprotokoll besitzt eine administrative Distanz, und diese ist in den Einträgen der Routingtabelle enthalten. Kleinere Werte werden bevorzugt, d.h., bei zwei Routen mit der gleichen Präfixlänge wird die mit der kleineren administrativen Distanz gewählt. Typische Beispiele finden Sie in Tabelle 1-8.

Tabelle 1-8: Administrative Distanzen verschiedener Protokolle

Routentyp	Administrative Distanz
Statisch	1
EIGRP	90
OSPF	110
RIP	120

Laut diesen Werten sind OSPF-Informationen gegenüber RIP zu bevorzugen. Bei der gleichen Präfixlänge würden Sie also OSPF-Informationen anstelle von RIP verwenden. Bietet RIP aber eine Route mit einer Präfixlänge von 24 an, während die Präfixlänge von OSPF nur 22 beträgt, dann wird die RIP-Information installiert und genutzt. In einer Routingtabelle stehen in eckigen Klammern Zahlen für die administrative Distanz:

RIP - 192.168.1.0 255.255.255.0 [120]

OSPF - 192.168.1.0 255.255.252.0 [110]

Beachten Sie, dass entsprechend der administrativen Distanz statische Routen allen erlernten Routen gegenüber bevorzugt werden und direkt verbundene Routen noch über statischen Routen stehen.

Metrik

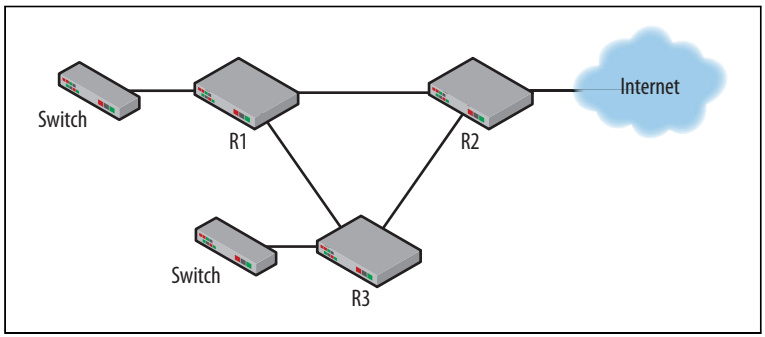
Die *Metrik* ist der letzte Vergleichswert bei Routeninformationen. Die Metrik wird verwendet, um Routen zu vergleichen, die über das gleiche Routingprotokoll erlernt wurden und die gleiche Präfixlänge aufweisen. Die Metrikwerte hängen vom Routingprotokoll ab – RIP verwendet die Anzahl der Hops, während OSPF eine Formel verwendet, um seine dimensionslose Metrik abzuleiten. Die Metrik ist zum direkten Vergleich der Informationen unterschiedlicher Protokolle nicht geeignet. Zum Beispiel werden zwei Pfade zum gleichen Ziel von einem Router über RIP-Pakete empfangen. Sie weisen also beide die gleiche administrative Distanz auf. Wenn wir annehmen, dass die Masken die gleiche Präfixlänge haben, ist die Metrik der entscheidende Faktor. Ein Pfad benötigt 4 Hops zum Ziel, während der andere nur 3 Hops benötigt. Ein Pfad ist also offensichtlich kürzer und wird daher in der Routingtabelle installiert. Die Routingtabelle enthält dann Einträge wie:

192.168.1.0 255.255.255.0 [120/3] via 192.168.1.254

Innerhalb der eckigen Klammern wird die Anzahl der Hops an die administrative Distanz angehängt.

Routingschleifen

Es gibt verschiedene Topologien, die sowohl Ethernet als auch IP Probleme bereiten. Eine Architektur mit Schleifen ist eine der größten Herausforderungen. Schicht-2-Protokolle wie Ethernet besitzen keinen Mechanismus zur Verarbeitung von Schleifen, weshalb Radia Perlman das Spanning Tree Protocol (STP) entwickelte. In der dritten Schicht gibt es ein gewisses Maß an Schutz, da IP ein Time-to-Live-Feld besitzt. Während sich Pakete durch die Schleifen-Topologie bewegen, dekrementiert jeder Router dieses Feld, bis es den Wert Null erreicht. An diesem Punkt angekommen, wird das Paket nicht mehr weitergesendet. Eine einfache Schleifen-Topologie ist in Abbildung 1-13 zu sehen.



◀ **Abbildung 1-13**
Routingschleife

Bei dieser Topologie würden die mit den Switches verbundenen Knoten R1 und R2 als Standard-Gateways verwendet. R1 und R2 würden wiederum R3 als »letzten Ausweg« nutzen, um an externe Ziele zu gelangen. Das Routing zwischen R1 und R2 kann über statische oder dynamische Routen erfolgen. Wie weiter oben beschrieben wurde, besteht das Problem mit statischen Routen darin, dass sie nicht auf veränderte Netzwerkbedingungen reagieren und keine Schleifen verarbeiten können. Jeder Fehler in der Konfiguration oder im Umgang mit bestimmten Arten von Fehlern führt dazu, dass die Pakete unendlich kreisen oder verloren gehen.

Doch Routingschleifen sind nicht immer schlecht. Wenn beispielsweise die Konnektivität der mit den Switches verbundenen Knoten als besonders kritisch erachtet wird, dann kann eine Routingschleife eingerichtet werden, um das Netzwerk sehr zuverlässig zu machen.

Die Links zwischen R1/R3 und R2/R3 können über sehr lange Strecken laufen, wie etwa die Verbindungen zu einem Service-Prodiver. Routing/Failover-Protokolle könnten genutzt werden, um diesen Satz redundanter Links zu pflegen, insbesondere wenn die Topologie komplexer ist als die in Abbildung 1-13. Routingschleifen können eingerichtet werden, um ein Load Balancing zwischen den Links zu ermöglichen. Protokolle wie das Hot Standby Routing Protocol (HSRP), das Virtual Router Redundancy Protocol (VRRP) und das Gateway Load Balancing Protocol (GLBP) wurden alle entworfen, um einen »Single Point of Failure« zu vermeiden und Traffic zwischen Links zu verteilen.

Abbildung 1-13 ist eine sehr einfache Form von Schleife, aber natürlich nicht die einzige Möglichkeit einer Schleifen-Topologie. Fehlerhafte Konfiguration oder verlorene Verbindungen können sehr leicht zu Schleifen führen, selbst wenn es keine physischen Schleifen gibt. Tatsächlich besitzen Netzwerke zwei Topologien: physisch und logisch. Die physische Topologie kann man nachvollziehen, indem man alle Kabel verfolgt oder die Aufkleber studiert. Die logische Topologie kann man nur verstehen, indem man die Konfigurationen untersucht und den Datenfluss verfolgt. Ein Beispiel dafür, dass physische und logische Topologie nicht identisch sind, zeigt Abbildung 1-14.

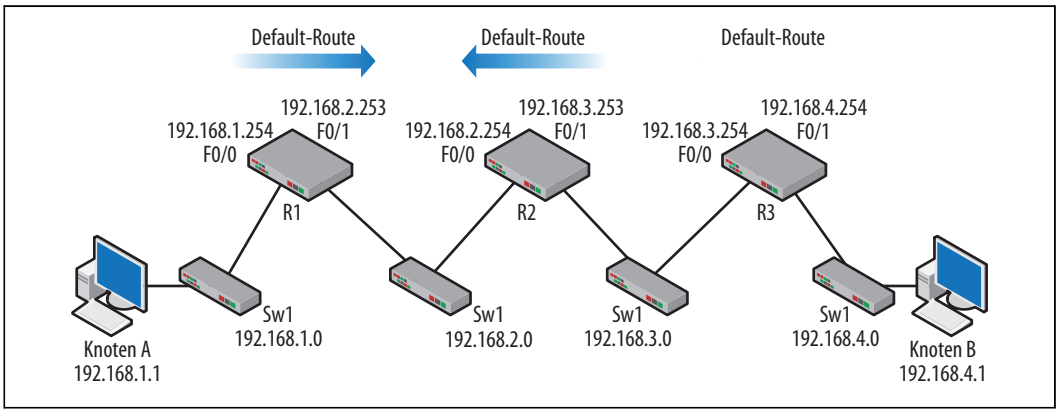


Abbildung 1-14 ▲ Physisch linear mit logischer Schleife

Bei unserer Diskussion zu statischen und Default-Routen zu Beginn dieses Kapitels wurde die Routingtabelle durch die Verwendung einer Default-Route auf R1 und R3 vereinfacht. Doch die Angabe einer Default-Route auf R2 hätte die Routingtabelle nicht vereinfacht. Wir wollen nun zeigen, warum der Eintrag einer Standard-Route auf R2 aus einem ganz anderen Grund keine gute Idee ist. Nehmen wir an, die Routingtabellen sind aufgebaut und die Stan-

dard-Routen sind zugewiesen (siehe Abbildung 1-14). R2 verwendet nun R1 als Default-Route.

Tabelle 1-9: Default-Routing in Routingschleife

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
S 0.0.0.0/0 via 192.168.2.254	S 0.0.0.0/0 via 192.168.2.253	S 0.0.0.0/0 via 192.168.3.253

Was passiert nun, wenn Knoten A ein Gerät anpingt, das nicht in dieser Netzwerkgruppe enthalten ist, etwa `192.168.5.1`? Der ICMP Echo-Request wird an das Default-Gateway von Knoten A gesendet (`192.168.1.254`), und R1 erkennt, dass er nicht weiß, wo das Ziel liegt. R1 sendet das Paket an sein »Gateway zum letzten Ausweg«: `192.168.2.254`. R2 geht seine Routingtabelle durch und erkennt, dass er das Ziel (`192.168.5.1`) ebenfalls nicht kennt. R2 besitzt ebenfalls einen »letzten Ausweg«, doch das Problem besteht darin, dass dieses Gateway R1 ist. Das Paket wird also umgehend an R1 zurückgeschickt. Ergo: logische Schleife. R1 empfängt das Paket, geht seine Routingtabelle durch, und dann geht die ganze Sache wieder von vorne los, bis das TTL-Feld des Pakets abläuft. Ob diese Konfiguration nun ein Versehen oder gewollt war, das Ergebnis ist das gleiche. Abbildung 1-15 zeigt ein ICMP-Paket (Internet Control Message Protocol), das aus einem TTL-Feld (Time-to-Live) resultiert, das auf 0 reduziert wird (wenn auch während einer anderen Kommunikation). ICMP ist dafür verantwortlich, Netzwerk-Hosts darüber zu informieren, wenn solche Probleme auftreten. Innerhalb des ICMP-Pakets wird das TTL-Feld auf 255 gesetzt, doch das gilt nicht für alle IP-Pakete. Jeder Router dekrementiert dieses Feld, während er das Paket weiterleitet.

Die in Abbildung 1-14 dargestellte Topologie ist isoliert. In der Praxis wäre sie mit der Außenwelt oder einer Reihe weiterer Router verbunden, die den Traffic letztendlich nach außen senden. Das Standard-Gateway und die Routingtabellen wären also richtig konfiguriert. Dennoch sollten Sie unsere Fähigkeit, die Dinge falsch einzurichten, nicht unterschätzen.

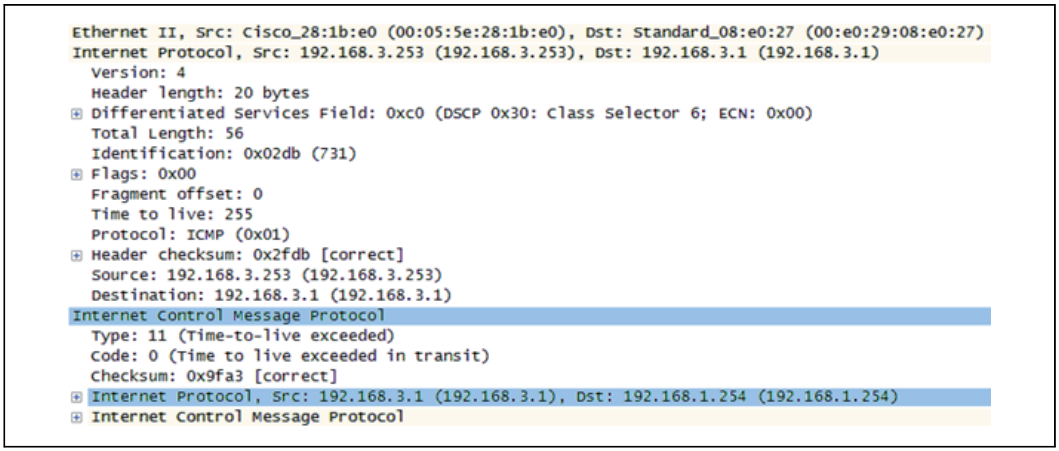


Abbildung 1-15 ▲
ICMP Time Exceeded

Es kommt vor, dass Fehler bei den Verbindungen zu Schleifen führen. Nehmen wir zum Beispiel an, dass in Abbildung 1-14 das R3-Interface, das mit `192.168.4.0` verbunden ist, heruntergefahren wird. Die Route wird aus der Routingtabelle von R3 entfernt. Doch die anderen Router in der Topologie glauben auch weiterhin, dass das Netzwerk `192.168.4.0` über R3 erreichbar ist. Die Frage lautet: Was macht R3, wenn Traffic für das `192.168.4.0`-Netzwerk hereinkommt?

Tabelle 1-10: Korrekte Routingtabellen – mal wieder

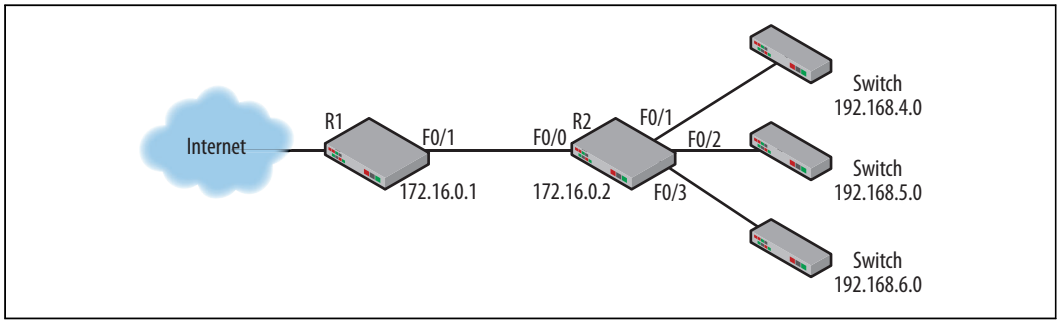
R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
S 0.0.0.0/0 via 192.168.2.254	S 192.168.1.0 via 192.168.2.253	S 0.0.0.0/0 via 192.168.3.253
	S 192.168.4.0 via 192.168.3.254	

Wie wir sehen können, empfängt R3 das Paket und sendet es, da es das Ziel nicht kennt, an sein »Gateway zum letzten Ausweg« (R2). Das liegt daran, dass beim Herunterfahren des Interfaces R3 das Netzwerk `192.168.4.0` aus seiner Routingtabelle entfernt, was zu einer weiteren Routingschleife führt. Wenn es also Zweifel an der Stabilität eines Netzwerks gibt oder wenn es komplexer wird, dann sollten Sie auf dynamisches Routing setzen.

Discard-Routing oder Null-Routing

Manchmal passen auch die besten Designs nicht zur jeweiligen Topologie. Wenn das passiert, können Versuche, das Netzwerk zu vereinfachen oder zu optimieren, zu echten Kopfschmerzen führen.

Beispielsweise wird häufig die Aggregation genutzt, um Routingtabellen zu verkleinern oder zu vereinfachen. Um eine Reihe von Routen zu aggregieren, muss die Zahl der zu aggregierenden Downstream-Routen auf Vielfachen der Basis 2 basieren. Wenn zur Aggregation von Routen genutzte Netzwerkmasken modifiziert werden, basieren diese Änderungen auf Vielfachen der Basis 2. Sehen wir uns ein Beispiel an: Nehmen wir an, der Netzwerkadministrator möchte die Routingtabellen für die kleine, aggregierte Topologie in Abbildung 1-16 aufräumen.



▲ **Abbildung 1-16**
Aggregierte Topologie

Die Routingtabellen für R1 und R2 sind in Tabelle 1-11 zu sehen. Bei diesem Beispiel kümmern wir uns nicht um die Außen-Konnektivität von R1.

Tabelle 1-11: Routingtabellen, aggregierte Topologie

R1	R2
C 172.16.0.0/16 F0/1	C 172.16.0.0/16 F0/0
S 192.168.4.0/24 via 172.16.0.2	C 192.168.4.0/24 F0/1
S 192.168.5.0/24 via 172.16.0.2	C 192.168.5.0/24 F0/2
S 192.168.6.0/24 via 172.16.0.2	C 192.168.6.0/24 F0/3
	S 0.0.0.0/0 via 172.16.0.1

Die Routingtabellen zeigen, dass R2 R1 als Default-Gateway verwendet und dass R1 verschiedene Netzwerke über R2 erreicht. Der Netzwerkadministrator sieht sich das an und entscheidet sich für eine Aggregation der beiden, um die Routingtabelle für R1 zu vereinfachen. Das wird durch eine Veränderung der Maske für die Downstream-Routen auf R1 erreicht.

Tabelle 1-12: Routingtabellen, aggregierte Topologie mit Admin-»Korrektur«

R1	R2
C 172.16.0.0/16 F0/1	C 172.16.0.0/16 F0/0
S 192.168.4.0/22 via 172.16.0.2	C 192.168.4.0/24 F0/1
	C 192.168.5.0/24 F0/2
	C 192.168.6.0/24 F0/3
	S 0.0.0.0/0 via 172.16.0.1

Der resultierende Eintrag in R1 umfasst nun die folgenden Adressen: `192.168.4.0` bis `192.168.7.255`. Doch was passiert, wenn eine Adresse wie `192.168.7.1` von außerhalb R1 angepingt wird? Der Traffic wird an R2 weitergeleitet, doch da die Route nicht Teil der Tabelle von R2 ist, verwendet er seine Standard-Route, um sie gleich wieder an R1 zurückzuschicken. Und wieder haben wir es mit einer Schleife zu tun. Eine Lösung dieses Problems kann darin bestehen, Null-Routen auf R2 einzutragen, um zu verhindern, dass er Traffic zurück an R1 schickt. Das kann für aggregierte Routen oder für kleinere Adressräume verwendet werden, d.h., Varianten dieses Befehls können auf jedem Router verwendet werden.

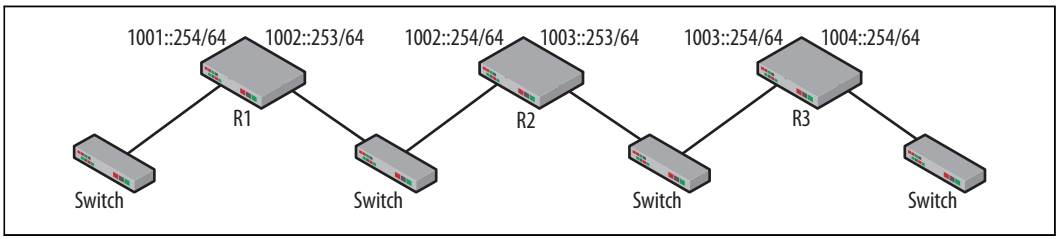
```
ip route 192.168.4.0 255.255.252.0 null0
```

Um zu verhindern, dass diese Route den gesamten Traffic stoppt, können Sie ihr eine höhere administrative Distanz zuweisen.

IPv6

Zwar ist IPv6 nicht der Schwerpunkt dieses Buches, doch es kann nicht schaden, einen Blick zu riskieren. Der schwierige Teil von IPv6 ist das Erlernen der Adressierung und der ganzen Begriffe. Schließlich müssen Sie Ihren Geist auf Werte vorbereiten, die so ganz anders aussehen. Doch aus Sicht der Routings sind viele Techniken gleich. Abbildung 1-17 zeigt die vorhin verwendete Topologie in ihrer IPv6-Variante. Das /64 ist die CIDR-Notation für die verwendeten Masken.

Abbildung 1-17 ▼
IPv6-Topologie



Um eine Topologie aufzubauen, ist eine Reihe von Änderungen an der Konfiguration aller Router notwendig. Wie man sehen kann, besitzt jedes Router-Interface eine IPv6-Adresse. Für R1 sind die IPv4- und IPv6-Befehle fast identisch:

```
ip address 192.168.1.254  255.255.255.0    ipv6 address
   1001::254/64
```

Der Hauptunterschied in der Struktur ist wieder die Adresse. Die Doppelpunkte der IPv6-Adresse unterdrücken lange, nur aus Nullen bestehende Strings. Das /64 ist ein CIDR-Kürzel (Classless Inter-Domain Routing) für die Maske. Das Routing wird mit zwei Befehlen eingerichtet: `ipv6 unicast-routing` und `ipv6 route`. Der zweite Befehl ist für statische Routen gedacht. Bei R1 werden Routen für die Netzwerke 1003::/64 und 1004::/64 benötigt.

```
ipv6 route 1003::/64 1002::254
ipv6 route 1004::/64 1002::254
```

Die Routingtabelle für IPv6-basierte Routen kann auf den ersten Blick verwirrend sein, doch wenn man die Sache aufdröselt, werden die Ähnlichkeiten deutlich. Abbildung 1-18 zeigt die Routingtabelle für R1. Beachten Sie die Verwendung direkt verbundener und statischer Routen. Eine Ergänzung ist der L- (Local) oder Link Local-Eintrag. Er verweist auf das Interface des Routers. Die Maske für diese Einträge ist /128, besteht also nur aus Einsen. Das entspricht dem IPv4-Host-Eintrag. FF00 ist der Multicast-Eintrag. Die eckigen Klammern der Einträge geben auch hier die administrative Distanz und die Metrik an.

▼ Abbildung 1-18
Die IPv6-Routingtabelle für R1

```
IPv6 Routing Table - Default - 7 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, M - MIPv6, R - RIP, I1 - ISIS L1
       I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary, D - EIGRP
       EX - EIGRP external
       O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
C   1001::/64 [0/0]
     via FastEthernet0/0, directly connected
L   1001::254/128 [0/0]
     via FastEthernet0/0, receive
C   1002::/64 [0/0]
     via FastEthernet0/1, directly connected
L   1002::253/128 [0/0]
     via FastEthernet0/1, receive
S   1003::/64 [1/0]
     via 1002::254
S   1004::/64 [1/0]
     via 1002::254
L   FF00::/8 [0/0]
     via Null0, receive
```

Lektüre

Die in diesem Kapitel behandelten Ideen werden in einer Reihe von RFCs und Standards beschrieben oder erwähnt, wenn man ein Protokoll liest. Beispielsweise behandeln die RFCs für RIP und OSPF verschiedene Routingaspekte und sind deshalb hier aufgeführt. Bei der Konfiguration von Netzwerkausrüstung fand ich zwei Dokumente immer hilfreich: die Befehlsreferenz und die Konfigurationsanleitung. Die Befehlsreferenz ist ein Muss, da sie die eigentlichen Befehle und die von ihnen genutzten Argumente aufführt. Allerdings sind sie nicht besonders hilfreich, wenn man die »bewährten Praktiken« kennenlernen will. Da kommt das Konfigurationshandbuch ins Spiel. Dieses Dokument erläutert, zusammen mit den Whitepapers des Herstellers, wo ein bestimmter Befehl sinnvollerweise verwendet wird oder wie man mit dem Aufbau eines Netzwerks beginnt. Doch während Sie versuchen die Dinge ans Laufen zu bekommen und Ihre Erfahrungen sammeln, macht letztlich nur der Versuch klug.

IEEE 802.1D: »Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges«
RFC 1102: »Policy Routing in Internet Protocols«
RFC 2328: »OSPF version 2«
RFC 2453: »RIP version 2«
RFC 3768: »Virtual Router Redundancy Protocol«

Zusammenfassung

In diesem Kapitel haben wir uns einige allgemeine Aspekte des Routings und Switchings angesehen. Beim Aufbau von Netzwerken ist es wichtig, die grundlegenden Konzepte wie statische, dynamische und Standard-Routen verstanden zu haben. Wenn das Netzwerk in Größe und Komplexität wächst, wird es wichtig, solide Topologie-Entscheidungen zu treffen und Routingprotokolle auswerten zu können. VLANs, Trunks, die Installation von Routen und Schleifenarchitekturen wurden ebenfalls behandelt. Die besten Netzwerkadministratoren verstehen nicht nur die von ihnen verwendeten Befehle, sondern wissen auch, warum sie diese Befehle verwenden und wie Entscheidungen im Netzwerk getroffen werden.

Fragen zum Kapitel

1. Bei der Zusammenschaltung und dem Betrieb von VLANs wird häufig die Bündelung (Trunk Lines) verwendet.
   a. WAHR
   b. FALSCH
2. Statische Routen werden von Hand eingetragen und besitzen eine niedrigere administrative Distanz als dynamische Routen.
   a. WAHR
   b. FALSCH
3. Dynamische Routen werden immer verwendet, wenn sich die Netzwerk-Topologie ändert.
   a. WAHR
   b. FALSCH
4. Wie lautet entsprechend der Bedeutung die richtige Reihenfolge für die Routenwahl?
   a. Administrative Distanz, Präfixlänge, Metrik
   b. Präfixlänge, administrative Distanz, Metrik
   c. Metrik, Präfixlänge, administrative Distanz
5. Der »nächste Hop«-Router muss ein Interface an einem mit Ihrem Netzwerk verbundenen Router sein.
   a. WAHR
   b. FALSCH
6. Direkt verbundene Routen werden installiert, sobald ein Interface »oben« ist.
   a. WAHR
   b. FALSCH
7. Ordnen Sie die folgenden Begriffe ihren Definitionen zu.

a. Metrik	1. Anzahl der Bits in der Maske
b. Administrative Distanz	2. Wert, der Informationen des gleichen Routingprotokolls vergleicht
c. Präfixlänge	3. Qualitätsvergleich zwischen Routingprotokollen

8. Welche Art von Routingprotokoll sind RIP und OSPF?
   a. Distanzvektor
   b. Host-basiert
   c. Hierarchisch
   d. Intern

9. Welche Adressen werden durch den folgenden Eintrag in der Routingtabelle erfasst: `172.31.32.0/19`?
10. Routingschleifen kommen nur in Netzwerken mit physischen Schleifen vor.
    a. WAHR
    b. FALSCH

Antworten

1. WAHR
2. WAHR
3. FALSCH
4. B
5. WAHR
6. WAHR
7. a) 2 b) 3 c) 1
8. D
9. 172.31.32.0–172.31.63.255
10. FALSCH

Laborübungen

Übung 1: Verbundene Switches und SATs

Material: Zwei Switches, zwei Computer

1. Verbinden Sie zwei Switches mit einem gekreutzen Kabel oder über einen Uplink.
2. Verbinden Sie die beiden Computer mit jeweils einem Switch.
3. Untersuchen Sie die SAT auf jedem Switch. Achten Sie auf die VLAN-, Port- und MAC-Adress-Listings. Praktischer Cisco-Befehl: `show mac-address-table`.
4. Ändern Sie die Lage der Computer, oder fügen Sie weitere Knoten hinzu.
5. Bevor Sie die Tabelle nach jedem Experiment untersuchen, sagen Sie den SAT-Inhalt voraus, und beschreiben Sie, warum das so ist.

Übung 2: Statische Routingtopologie

Material: Drei Router, zwei Computer

1. Bauen Sie die Topologie aus Abbildung 1-7 auf. Hinweis: Die Topologie kann auf zwei Router mit den gleichen Anforderungen reduziert werden.
2. Weisen Sie den Router-Interfaces und den Computern IP-Adressen zu.
3. Untersuchen Sie die Routingtabellen der Router, sobald die Interfaces oben sind. Praktischer Cisco-Befehl: `show ip route`.
4. Experimentieren Sie mit PING. Welche Ziele sind erreichbar und welche nicht?
5. Arbeiten Sie sich von links nach rechts durch, und beginnen Sie damit, statische Routen einzutragen, um Konnektivitätsprobleme zu beheben. Praktischer Cisco-Befehl: `ip route destination network destination mask forwarding router interface`.
6. Wenn alle Ziele von allen Schnittstellen aus angepingt werden können, sind Sie fertig.

Übung 3: Umwandlung in Default-Routen

Material: Drei Router, zwei Computer, Wireshark

1. Nutzen Sie die Topologie des vorangegangenen Experiments, und wandeln Sie die statischen Routen auf R1 und R3 in Default-Routen um. Hinweis: Das kann verwirrend sein, wenn nur zwei Router verwendet werden, da es keinen klaren Grund für die Wahl der Default-Route gibt.
2. Untersuchen Sie die Routingtabellen aller Router. Wählen Sie einige Ziele aus, und verarbeiten Sie die Routingtabellen von Hand. Überprüfen Sie, ob Sie dem Prozess Schritt für Schritt folgen können.
3. Experimentieren Sie nun mit den Captures selbst. Beginnen Sie mit einem Computer oder Interface und gehen Sie davon aus, dass die ARP-Tabellen leer sind. Versuchen Sie nun, jedes Paket zu erläutern, das durch einen PING auf eine mindestens einen Hop entfernte IP-Adresse generiert wird.
4. Schließen Sie den PING ab, und untersuchen Sie die Captures, um die richtige Antwort zu bestimmen. Waren Ihre Antworten richtig? Wenn nicht, warum nicht?

Übung 4: Routingschleife

Material: Drei Router, zwei Computer, Wireshark

1. Verwenden Sie die gleiche Topologie wie oben, und wandeln Sie die Routingtabelle auf R2 in Default-Routen um.
2. Welche Adressen können angepingt werden und welche nicht?
3. Was passiert in der Shell, wenn Sie eine in der Topologie nicht vorhandene Adresse anpingen?
4. Starten Sie Wireshark, und untersuchen Sie den Netzwerk-Traffic für den nach außen führenden PING.
5. Was passiert mit dem IP-TTL-Feld?
6. Wo war die Schleife, und was hat sie verursacht?
7. Wie sieht der daraufhin generierte ICMP-Traffic aus?

Übung 5: Null-Route

Material: Drei Router, zwei Computer, Wireshark

1. Installieren Sie in obiger Topologie Null-Routen, um die Routingschleife zu beheben. Denken Sie an das Null-Argument des ip route-Befehls zurück.
2. Löst das Ihr Konnektivitätsproblem, oder versteckt das nur die Schwierigkeiten?

KAPITEL 2

Host-Routing

In diesem Kapitel:

- Der Entscheidungsprozess
- Host-Routingtabellen
- Adressierung
- Paket-Tracking
- Lektüre
- Zusammenfassung
- Fragen
- Antworten
- Laborübungen

»Ganz schön viel Aufwand, nur um ein Paket von einer Seite des Raums zur anderen zu bewegen.«

Ein anonymer Student

Kapitel 1 hat verschiedene Aspekte des Forwardings von Daten über ein Netzwerk beleuchtet. Die Kommunikation läuft von einem Host (üblicherweise einem Server irgendeiner Art) zum anderen und dann wieder zurück. Die Switch-Quelladresse und die Router-Routingtabellen sind für diesen Prozess unabdingbar. Doch unabhängig vom Zweck der Datenübertragung müssen verschiedene Operationen durchgeführt werden, bevor Pakete ins Netzwerk gesendet werden können. Den Anfang macht da die Routingtabelle des Hosts. Damit eng verknüpft sind Ideen wie Maskierung, Adressauflösung und Standard-Gateways.

Der Entscheidungsprozess

Von dem Moment an, in dem ein Quell-Host ein Datensegment für die Übertragung erzeugt, beginnt ein Prozess, der schließlich in einem zu übertragenden Ethernet-Frame endet. Von der Anwendungsschicht werden die Daten weiter nach unten durchgereicht und in einer Reihe von Headern gekapselt, bis das untere Ende des Protokollstacks erreicht ist. Zum Beispiel wird beim Zugriff auf eine Webseite das HTTP-Protokoll (Hypertext Transfer Protocol) genutzt, um Informationen zwischen Webserver und Host auszutauschen. HTTP verwendet TCP (Transmission Control Protocol) an Schicht 4, gefolgt von IP und dann Ethernet oder 802.11. Wenn wir von Ethernet ausgehen, sehen die gekapselten Daten aus wie in Abbildung 2-1.

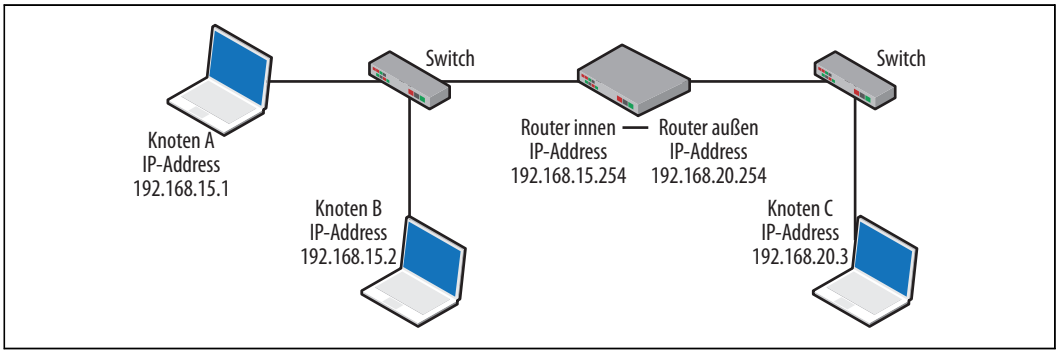
```
Ethernet II, Src: WesternD_89:ba:fa (00:00:c0:89:ba:fa), Dst: Cisco_2c:0c:80 (00:11:21:2c:0c:80)
Internet Protocol, Src: 192.168.1.1 (192.168.1.1), Dst: 192.168.1.254 (192.168.1.254)
Transmission Control Protocol, Src Port: cma (1050), Dst Port: http (80), Seq: 1, Ack: 1, Len: 405
Hypertext Transfer Protocol
```

Abbildung 2-1 ▲
Kapselung

Wenn sich die Kapselung dem Abschluss nähert, müssen die MAC-Adressen von Quelle und Ziel eingefügt werden. Das gilt auch für die Adressierung in anderen Schichten, doch die IP-Adressen und Portnummern sind eindeutig, da der Host mit einem IP-basierten Server über bekannte Ports kommuniziert. Tatsächlich ist auch die Quell-MAC-Adresse eindeutig, da der Host den Frame erzeugt. Die Frage lautet also: Welche Adresse soll im Ziel-MAC-Adressfeld des Ethernet-Frames stehen?

Wenn der Host die korrekte Ziel-IP-Adresse ermitteln kann, liefert ARP (das Address Resolution Protocol) die Antwort. Anders ausgedrückt: Wir müssen nach der richtigen MAC-Adresse fragen. Die Antwort hängt davon ab, ob das Ziel im gleichen Netzwerk liegt wie die Quelle. Die folgenden Seiten enthalten eine Reihe von Beispielen. Die Topologie in Abbildung 2-2 enthält die beiden Netzwerke `192.168.15.0` und `192.168.20.0`, die durch einen Router getrennt sind. Die Knoten A und B liegen im gleichen Netzwerk, während Knoten C in einem anderen Netzwerk liegt.

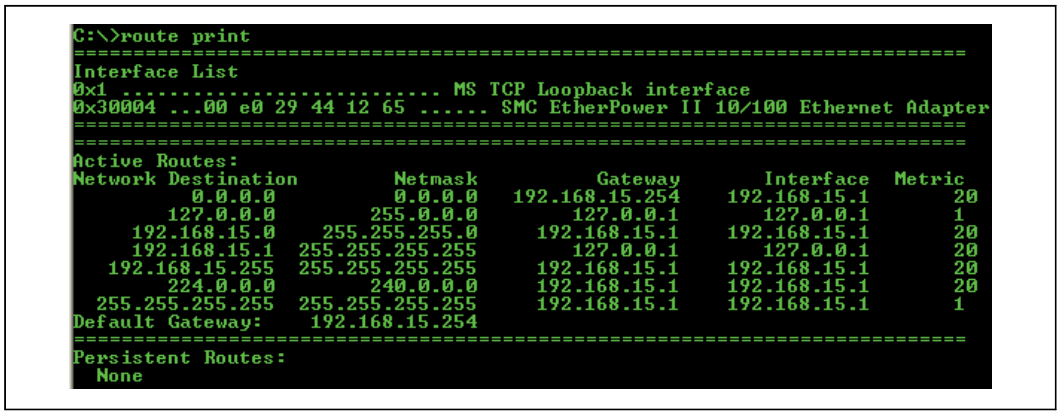
Abbildung 2-2 ▼
Kleine Topologie



Daten an Knoten innerhalb des gleichen Netzwerks zu senden ist sehr simpel. Die Übertragung läuft einfach von der Quell-IP- und -MAC-Adresse zur Ziel-IP- und -MAC-Adresse. Da wir ARP verwenden, um die Ziel-MAC-Adresse zu ermitteln, und da Switches die MAC-Adresstabelle (also die Quelladresstabelle) zur Weiterleitung von Ethernet-Frames verwenden, sprechen wir bei Knoten im gleichen Netzwerk von Forwarding-Entscheidungen der zweiten Schicht.

Was die Dinge etwas erschwert, ist die Tatsache, dass Daten aus dem anderen Netzwerk irgendwie zum Router gelangen müssen. ARP wird nicht verwendet, um MAC-Adressen in anderen Netzwerken aufzulösen. Es zeigt sich, dass beim Senden von Daten aus dem Netzwerk der Host die *MAC-Adresse des Routers* in den Ethernet-Frame einfügt. Diese Entscheidung wird getroffen, wenn der Host seine lokale Routingtabelle verarbeitet, die auch Host-Routingtabelle genannt wird. Die Routingtabelle für Knoten A ist in Abbildung 2-3 zu sehen. Die Ausgabe wurde auf einem Windows-PC mithilfe des Befehls `route print` in der Eingabeaufforderung erzeugt.

▼ **Abbildung 2-3**
Routingtabelle für Knoten A



Die meisten Betriebssysteme liefern vergleichbare Informationen. Es gibt fünf Spalten, und bei der Verarbeitung dieser Tabelle beginnen wir mit dem untersten Eintrag. Die Arbeit beginnt mit den Spalten 1 und 2. Zuallererst muss ein Knoten bestimmen, ob das Ziel im gleichen Netzwerk liegt oder nicht. Das Ergebnis bestimmt die nachfolgenden Schritte und den Ethernet-Header. Der zur Bestimmung des Netzwerks verwendete Mechanismus wird ANDing genannt. Das Begleitbuch zu diesem Titel, *Praxiskurs Netzwerkgrundlagen*, widmet der Maskierung ein ganzes Kapitel, aber wir wollen die Maskierung hier nur kurz zusammenfassen.

Die Netzwerkmaske wird verwendet, um die Netzwerk-ID der fraglichen IP-Adresse über ein logisches UND zu bestimmen. Ein Binärwert, der über ein logisches UND mit 0 verknüpft wird, liefert im Ergebnis eine 0. Die UND-Verknüpfung mit einer 1 behält den Originalwert bei. Wandelt man die IP-Adresse `172.16.49.67` und die Maske `255.255.224` in Binärwerte um, erhält man Folgendes:

10101100 . 00010000 . 00110001 . 01000011

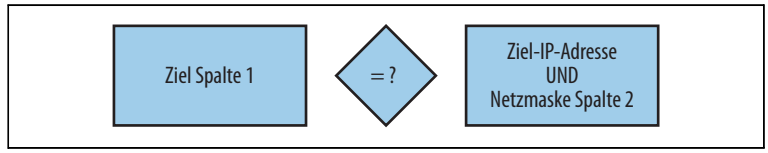
11111111 . 11111111 . 11100000 . 00000000

Eine bitweise UND-Verknüpfung ergibt:

10101100 . 00010000 . 00100000 . 00000000

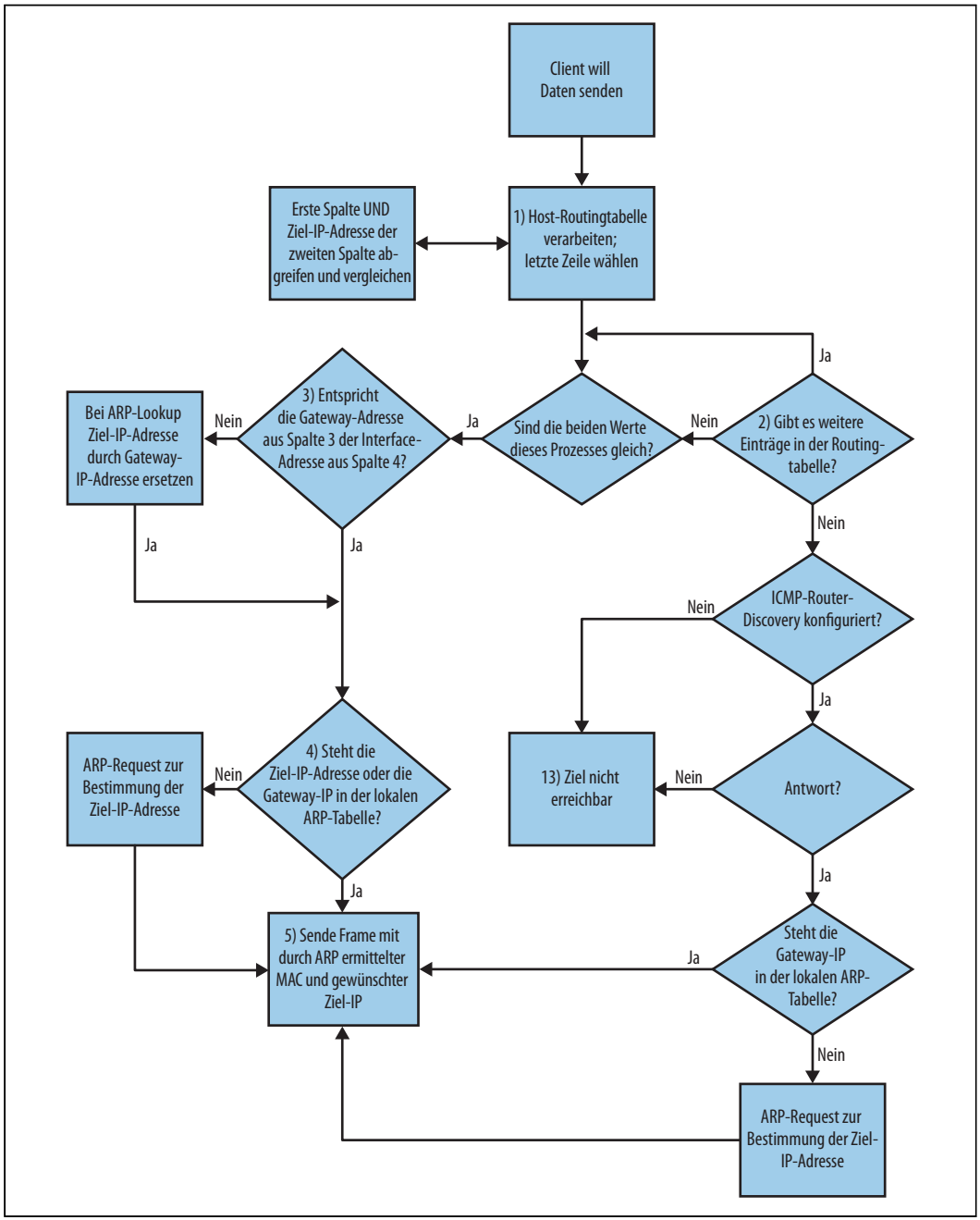
Die Umwandlung in Dezimalzahlen offenbart, dass diese IP-Adresse im Netzwerk `172.16.32.0` liegt. Es ist wichtig, daran zu denken, dass der Host versucht, sowohl die Quell- als auch die Zielnetzwerke zu bestimmen. Zuerst wird das Ziel aus der ersten Spalte genommen. Im zweiten Schritt erfolgt die UND-Verknüpfung der Zieladresse mit der Maske in der zweiten Spalte. Die grundlegende Frage lautet: Sind die resultierenden Werte identisch? Wenn nicht, versuchen Sie es erneut.

Abbildung 2-4 ▶
Die grundlegende Frage



In den folgenden Beispielen versucht Knoten A zuerst mit Knoten B im gleichen Netzwerk und dann mit Knoten C in einem anderen Netzwerk zu kommunizieren. Das Flussdiagramm in Abbildung 2-5 zeigt den Entscheidungsbaum, den ein Host unabhängig vom Zielnetzwerk durchläuft, wenn er versucht, einen anderen Knoten zu kontaktieren. Verschiedene eingebundene Prozesse sind nötig, um zu erläutern, wie das funktioniert. Unabhängig vom Ziel beginnen wir mit dem Parsing der Host-Routingtabelle. Die Schritte 1 und 2 in Abbildung 2-5 gehen jede Zeile der Tabelle durch, bis ein Treffer erkannt wird oder keine weiteren Einträge mehr vorhanden sind.

▼ **Abbildung 2-5**
Entscheidungsflussdiagramm



1. Fall: Das Ziel liegt im gleichen Netzwerk wie die Quelle

Bei diesem Beispiel gehen wir davon aus, dass Knoten A versucht, Knoten B anzupingen. Entsprechend der Tabelle in Abbildung 2-3 könnte die erste Entscheidung wie folgt aussehen:

		192.168.15.2
255.255.255.255	verglichen mit	UND
		255.255.255.255
		192.168.15.2

Da diese nicht gleich sind, wird die zweite Zeile verarbeitet:

		192.168.15.2
224.0.0.0	verglichen mit	UND
		240.0.0.0
		192.0.0.0

Auch diese sind nicht gleich.

In diesem Fall wird die Prozedur bis zur fünften Zeile von unten wiederholt:

		192.168.15.2
192.168.15.0	verglichen mit	UND
		255.255.255.0
		192.168.15.0

Da die Werte nun übereinstimmen, wird es Zeit, auf die linke Seite des Flussdiagramms zu Schritt 3 zu wechseln. An dieser Stelle muss das Interface bestimmt werden, das als Gateway verwendet werden soll. Der dritte Schritt überprüft, ob die IP-Adressen in der dritten und vierten Spalte identisch sind. Die in der vierten Spalte stehenden Interfaces sind einfach die IP-Adressen, die den Netzwerkkarten des Hosts zugeordnet sind. Die in Spalte 3 angegebenen Gateways sind entweder die gleichen Netzwerk-Interfaces oder mit dem Netzwerk verbundene Router. Entsprechend Abbildung 2-3 lautet die IP-Adresse des Interfaces entweder `192.168.15.1` oder loopback. Zeile 5 der Routingtabelle zeigt, dass sowohl die Interface-IP-Adresse für dieses Ziel als auch die Gateway-IP-Adresse `192.168.15.1` lautet. Das bedeutet, *dass dieses Interface das Gateway ist*. Mit anderen Worten: Es besteht keine Notwendigkeit, die Daten an das Default-Gateway zu senden, da das Ziel im gleichen Netzwerk liegt und direkt kontaktiert werden kann.

Nachdem feststeht, dass das Ziel im gleichen Netzwerk liegt, beginnt der Host mit dem Aufbau des Ethernet-Frames. Dazu benötigt er aber die MAC-Adresse des Ziels. Im vierten Schritt wird zu ARP gewechselt. Die Aufgabe von ARP ist die Ermittlung der MAC-Adresse des Ziel-Hosts. Diese Information wird lokal in der ARP-Tabelle des Hosts gespeichert. Zuerst wird die ARP-Tabelle untersucht, um herauszufinden, ob der Host die richtige MAC-Adresse bereits kennt. Wird kein Eintrag für die Ziel-IP-Adresse gefunden, muss ein ARP-Request gesendet werden. Sobald die Adresse erlernt wurde, kann das Ethernet-Frame konstruiert und an das Ziel gesendet werden.

2. Fall: Das Ziel liegt in einem anderen Netzwerk als die Quelle

Der zweite Fall beginnt mit dem gleichen Schritt wie der erste (d.h. mit der Verarbeitung der Host-Routingtabelle). Der einzige Unterschied ist die IP-Adresse des Ziels. Eine andere Zeile der Host-Routingtabelle liefert uns die Antwort. Diesmal ist die IP-Adresse von Knoten C das Ziel: `192.168.20.1`. Die erste Entscheidung der UND-Verknüpfung würde wie folgt aussehen:

	255.255.255.255		192.168.20.1
	UND	verglichen mit	UND
	255.255.255.255		255.255.255.255
Ergebnis	255.255.255.255		192.168.20.1

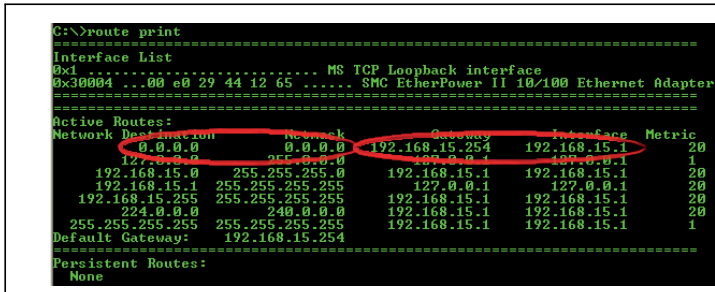
Für dieses Ziel würde der Prozess bis zur letzten Zeile der Routingtabelle weiterlaufen. Diese Zeile ist etwas Besonderes, da sie das Standard-Gateway darstellt, was durch das `0.0.0.0` für das Zielnetzwerk und die Netzwerkmaske angezeigt wird. Aufgrund des Ergebnisses der UND-Verknüpfung wird sie als »match all«-Zeile bezeichnet.

	0.0.0.0		192.168.20.1
	UND	verglichen mit	UND
	0.0.0.0		0.0.0.0
Ergebnis	0.0.0.0		0.0.0.0

Unabhängig vom Ziel sind die Ergebnisse der UND-Verknüpfung immer gleich. Wie im ersten Fall ist es nun Zeit für Schritt 3. Diesmal sind die Werte in der dritten und vierten Spalte unterschiedlich. Sind die beiden Werte nicht gleich, erkennt der Quell-Host, dass

das Ziel in einem anderen Netzwerk liegt, d.h., dass die Daten an das Standard-Gateway gesendet werden müssen, um das Ziel erreichen zu können. Die Zeile mit dem Standard-Gateway ist in Abbildung 2-6 eingekreist.

Abbildung 2-6 ▶
Standard-Gateway-Felder

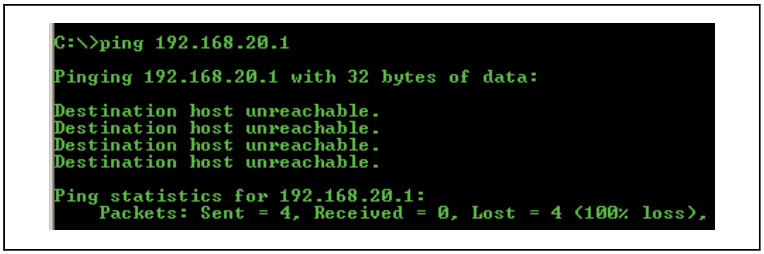


Zurück zum Flussdiagramm und zu Schritt 3: Ist das Gateway ein anderes als das Interface, muss der Knoten die Gateway-Adresse über ARP erfragen. Das liegt daran, dass der Frame aus dem Netzwerk herausgesendet werden muss. Die einzige Möglichkeit besteht darin, das für Knoten C gedachte IP-Paket zu nehmen und in einen Ethernet-Frame zu packen, der an den Router gesendet wird. Die Überprüfung der ARP-Tabelle bzw. ein ARPing von `192.168.15.254` liefert die MAC-Adresse des Standard-Gateways, und der Frame kann erzeugt werden.

Tipp Verwechseln Sie den Eintrag 0.0.0.0 für das Standard-Gateway nicht mit der IP-Adresse 0.0.0.0.

Was passiert, wenn das Standard-Gateway nicht bekannt ist?

Ein Teil des Flussdiagramms in Abbildung 2-5 widmet sich dem schwierigen Problem, dass es kein (oder ein falsches) Standard-Gateway gibt. Nutzt ein Host DHCP, ist es wahrscheinlich, dass eine Gateway-Adresse (zusammen mit einer IP-Adresse, Netzwerkmaske und DNS-Adresse) bereitgestellt wurde, aber eine Garantie gibt es dafür nicht. Das Gleiche gilt auch für statisch konfigurierte Hosts. Ein Standard-Gateway ist korrekt, wenn die IP-Adresse im gleichen Netzwerk liegt wie der Quell-Host. In Abbildung 2-6 liegen die Adressen `192.168.15.1` und `192.168.15.254` im gleichen Netzwerk. Fehlt der Gateway-Eintrag (`192.168.15.254`) oder ist er falsch, erscheint in der Kommandozeile die Ausgabe aus Abbildung 2-7.



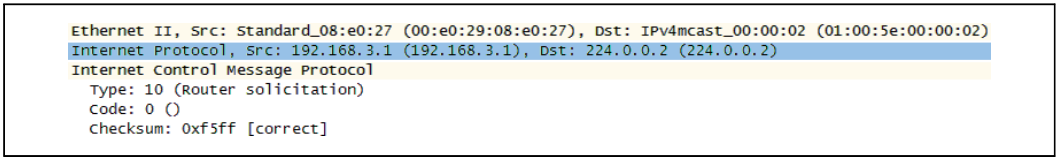
◀ **Abbildung 2-7**
Ausgabe bei einem fehlenden Standard-Gateway

Bei dieser Ausgabe sind eine Reihe von Dingen zu beachten:

- Aufgrund des Problems wurden keine Pakete erzeugt. Diese Nachricht wird vom lokalen Betriebssystem des Quell-Hosts generiert.
- Das wird häufig mit der »ICMP destination unreachable«-Meldung verwechselt, doch diese Meldungen sind NICHT identisch. Zwar werden die gleichen Begriffe verwendet, wenn die Ausgabe das Ergebnis einer ICMP-Meldung ist, doch die Quell-IP-Adresse ist nicht enthalten.

Das Problem eines fehlenden Standard-Gateways lässt sich leicht beheben, indem man z.B. die Konfiguration des DHCP-Servers oder die statische Konfiguration korrigiert. Aber auch ICMP-Router-Solicitation- und ICMP-Router-Advertisement-Nachrichten können verwendet werden. Diese Methode läuft standardmäßig üblicherweise nicht und muss sowohl auf dem Router als auch auf dem Host konfiguriert werden. Während sie bei kabelbasierten LANs nicht weiter verbreitet sind, nutzen Anwendungen wie Mobile-IP und einige Teile der Mobilfunk-Infrastruktur die ICMP-Nachrichten. Eine Router-Solicitation (Anforderung) ist in Abbildung 2-8 zu sehen. Beachten Sie, dass die Anforderung an die »Alle-Router«-Multicast-Adresse `224.0.0.2` gesendet wird.

▼ **Abbildung 2-8**
ICMP-Router-Solicitation



Entsprechend dem Flussdiagramm landet der Host beim Fehlen eines Standard-Gateways und ohne eine Host-Konfiguration für ICMP-Router-Solicitations (und Routern ohne Advertisements) beim unglückseligen Schritt 13 – Ziel nicht erreichbar (»destination unreachable«). Eine Host-Routingtabelle ohne Standard-Gateway ist in Abbildung 2-9 zu sehen.

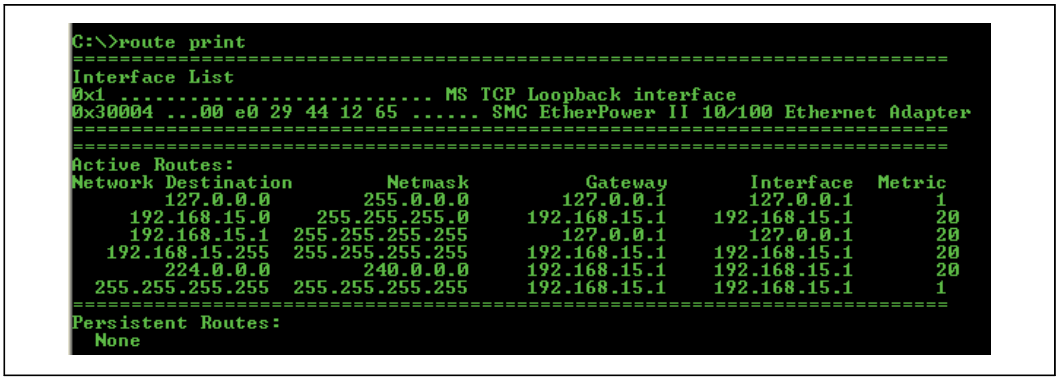


Abbildung 2-9 ▲
Routingtabelle ohne Standard-Gateway

Host-Routingtabellen

Zurück zur Host-Routingtabelle in Abbildung 2-9: Während die Routingtabelle verarbeitet wird, können Einträge auftauchen, die den IP zugeordneten Spezialadressen entsprechen. Diese Adresstypen, -werte und die Zeile der Routingtabelle für Knoten A sind nachfolgend zu sehen. Die Zeilen sind von oben nach unten durchnummeriert, denken Sie aber daran, dass die Host-Routingtabelle von unten nach oben verarbeitet wird.

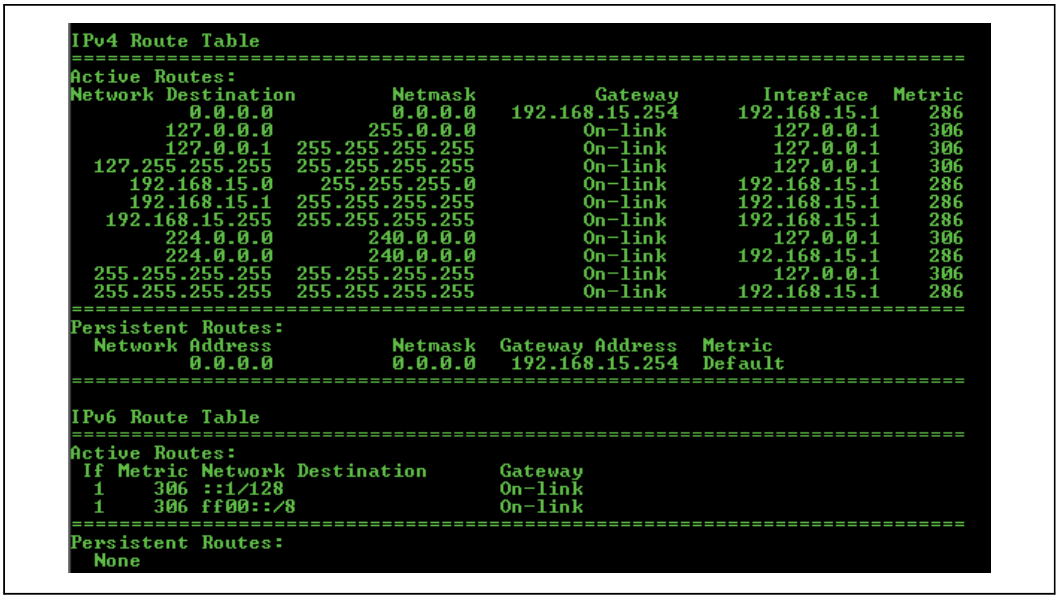
Tabelle 2-1: Spezielle Adressen

Aufgabe	Adresse	Zeile in Routingtabelle
Loopback	127.0.0.0	1
Netzwerk-ID	192.168.15.0	2
Host-IP-Adresse	192.168.15.1	3
Gerichteter Broadcast	192.168.15.255	4
Multicast	224.0.0.0	5
Eingeschränkter Broadcast	255.255.255.255	6

Tipp

Eine Adresse, die in vielen Netzwerken regelmäßig auftaucht, in Abbildung 2-9 aber fehlt, ist 169.254.0.0. Diese Adresse stammt vom IETF Zero Configuration-Standard und taucht üblicherweise auf, wenn der Host keine Adresse per DHCP oder eine statische Konfiguration erhalten hat.

Wie vorhin erwähnt wurde, sind die Einträge in der Host-Routingtabelle unabhängig vom Betriebssystem identisch, doch es gibt auch einige erwähnenswerte Unterschiede. Abbildung 2-10 zeigt die Routingtabelle eines Windows 7-PC, der in gleicher Weise konfiguriert wurde.



▲ **Abbildung 2-10**
Windows 7-Routingtabelle

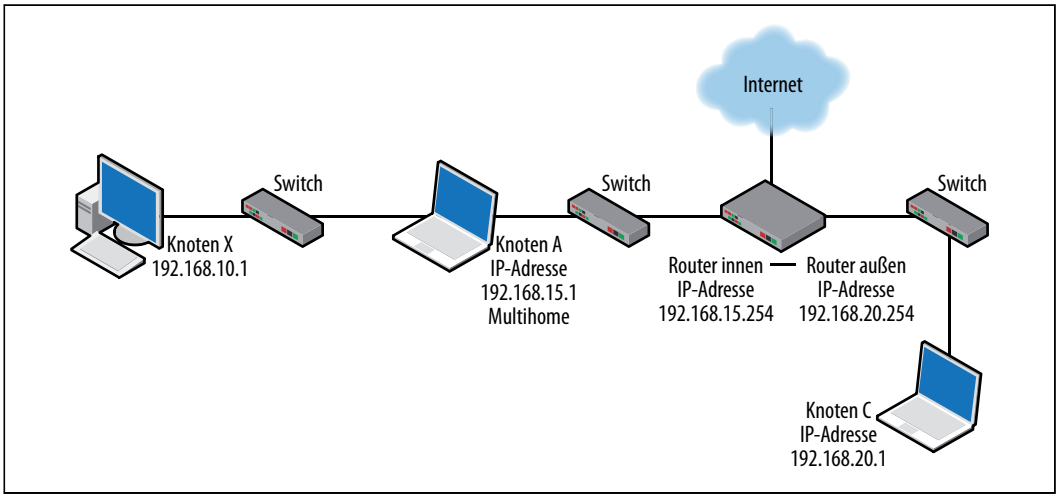
Der größte Unterschied ist die dritte Spalte. Sind die Gateway- und Interface-IP-Adressen identisch, dann liegt das Ziel im gleichen Netzwerk. Man kann das auch so sehen, dass das Ziel über eine Netzwerkschnittstelle des Quell-Hosts erreichbar ist. Das wird im deutschen Windows mit »auf Verbindung« gekennzeichnet. IPv6 bezeichnet diesen Typ als »Link-Local«. Es gibt auch eine Änderung in der fünften Spalte, die die Metrik beschreibt. Bei Hosts ist die Metrik üblicherweise an die Geschwindigkeit des Links gebunden. Hosts ziehen schnellere Verbindungen vor. Erinnern Sie sich an unsere Betrachtung aus Kapitel 1 bezüglich der Metrik. Die Metrik hat für Router eine ganz andere Bedeutung.

Der letzte wichtige Punkt im Bezug auf Host-Routingtabellen ist, dass die Beispiele bisher nur Single-Home-Hosts enthalten, d.h. einen Host mit einer einzelnen Netzwerkschnittstelle wie Knoten A, B und C in Abbildung 2-2. Dual- oder Multihomed-Computer besitzen zwei oder mehr Schnittstellen. Bei Multihome-Hosts erhöht sich entsprechend die Anzahl der Einträge in der Routingtabelle. Jede Schnittstelle besitzt einen eigenen Satz von Einträgen, wie etwa die sechs vorhin beschriebenen. Ein Beispiel für einen Multihome-Host ist ein Laptop, der verkabelt und dessen WiFi-Schnittstelle immer noch aktiv ist.

Bei Single-Home-Hosts ist die Gateway-Adresse einfach festzulegen. Der Host sendet seinen Traffic entweder direkt ins eigene Netzwerk oder an das oben beschriebene Standard-Gateway. Doch wenn das

Internet Connection Sharing (ICS) verwendet wird, um ein Interface in zwei Netzwerken unterzubringen, sieht die Routingtabelle etwas anders aus. Diese Topologie ist in Abbildung 2-11 zu sehen.

Abbildung 2-11 ▼
Knoten A ist ein Multihome-Host.



Sieht man sich die Routingtabelle für Knoten A noch mal an, erkennt man, dass sich nicht nur die Zahl der Einträge in der Routingtabelle verdoppelt hat, sondern dass es auch zwei Standard-Gateways gibt. Es gibt auch zwei persistente Routen, die die Netzwerke anzeigen, zu denen der Host eine direkte Verbindung hat.

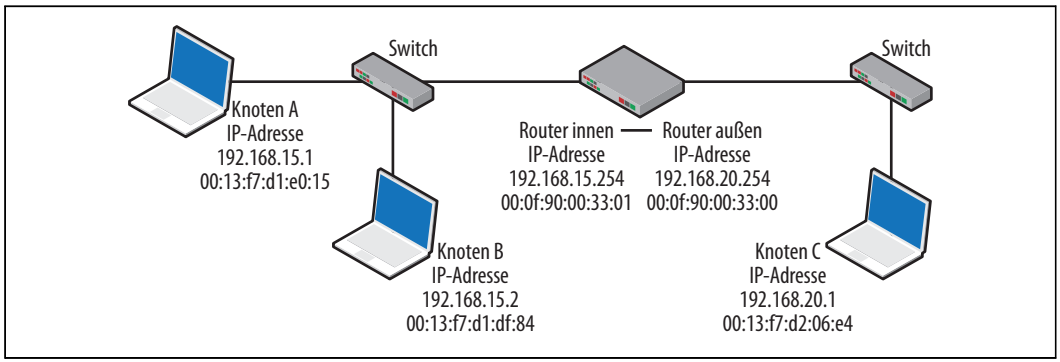
Abbildung 2-12 ▼
Routingtabelle bei Multihome-Host

```
IPv4 Route Table
===========================================================================
Active Routes:
Network Destination        Netmask          Gateway       Interface  Metric
          0.0.0.0          0.0.0.0   192.168.15.254    192.168.15.1     286
          0.0.0.0          0.0.0.0   192.168.10.254    192.168.10.1     281
        127.0.0.0        255.0.0.0         On-link         127.0.0.1     306
        127.0.0.1  255.255.255.255         On-link         127.0.0.1     306
  127.255.255.255  255.255.255.255         On-link         127.0.0.1     306
     192.168.10.0    255.255.255.0         On-link    192.168.10.1     281
     192.168.10.1  255.255.255.255         On-link    192.168.10.1     281
   192.168.10.255  255.255.255.255         On-link    192.168.10.1     281
     192.168.15.0    255.255.255.0         On-link    192.168.15.1     286
     192.168.15.1  255.255.255.255         On-link    192.168.15.1     286
   192.168.15.255  255.255.255.255         On-link    192.168.15.1     286
        224.0.0.0        240.0.0.0         On-link         127.0.0.1     306
        224.0.0.0        240.0.0.0         On-link    192.168.15.1     286
        224.0.0.0        240.0.0.0         On-link    192.168.10.1     281
  255.255.255.255  255.255.255.255         On-link         127.0.0.1     306
  255.255.255.255  255.255.255.255         On-link    192.168.15.1     286
  255.255.255.255  255.255.255.255         On-link    192.168.10.1     281
===========================================================================
Persistent Routes:
  Network Address          Netmask  Gateway Address  Metric
          0.0.0.0          0.0.0.0   192.168.15.254  Default
          0.0.0.0          0.0.0.0   192.168.10.254  Default
===========================================================================
```

Zwei Gateways hören sich nach einer guten Idee an, doch nur einer der Einträge (typischerweise der untere) wird verwendet. In diesem Fall würde der Host niemals Daten an `192.168.15.254` senden. Es könnte der Fall vorliegen, dass der untere Eintrag (`192.168.10.254`) nicht präferiert wird und die Daten so keinen Weg zum Ziel finden. Abbildung 2-11 enthält einen möglichen Weg ins Internet. Daten an das `192.168.10.0`-Netzwerk zu senden, endet in einer Sackgasse. Metriken können verwendet werden, um den verwendeten Eintrag zu beeinflussen, doch das ist nicht üblich.

Adressierung

In beiden Beispielen (Knoten A pingt Knoten B an, und Knoten A pingt Knoten C an) liefert der ARP-Prozess Einblick in die verwendete Adressierung. In diesem Abschnitt sehen wir uns die Pakete genauer an, die zwischen den Knoten hin und her laufen, sowie die in den Headern enthaltene Adressierung. Abbildung 2-13 zeigt die Original-Topologie, enthält aber auch die MAC-Adressen der Geräte.



▲ **Abbildung 2-13**
Topologie mit MAC-Adressen

Im ersten Fall spricht Knoten A per Ping (ICMP echo request) Knoten B an. Dieser Austausch benötigt die Router überhaupt nicht, und alle Pakete/Frames enthalten nur die IP- und MAC-Adressen der beiden fraglichen Knoten. Das folgende Paket umfasst auch den ersten zwischen den beiden gesendeten ICMP Echo Request. Dieses Paket wurde auf Knoten A festgehalten.

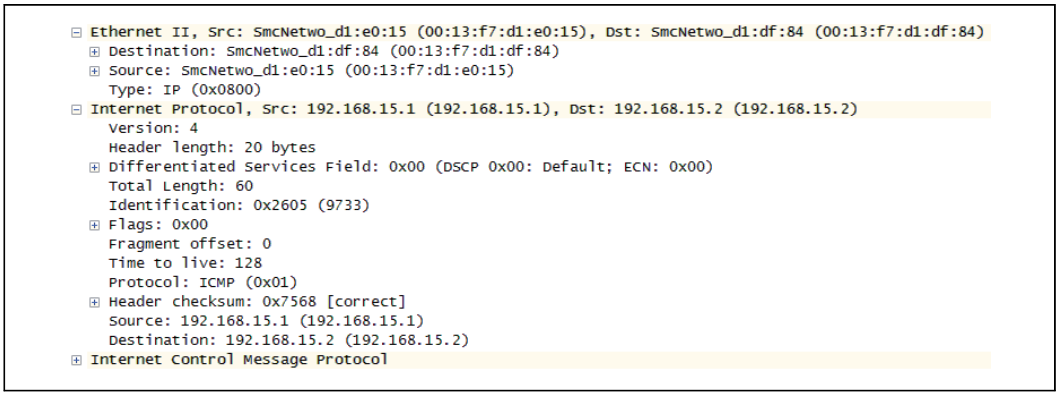


Abbildung 2-14 ▲
Im 1. Fall verwendete IP- und MAC-Adressen

Bei der Kommunikation mit einem Knoten außerhalb des Netzwerks (Knoten A pingt Knoten C an), wird ein Ethernet-Frame erzeugt und an den Router gesendet statt an den Zielknoten. Die IP-Adressen sind aber immer noch die von Knoten A und C. Unter normalen Umständen schreibt der Router die IP-Header nicht um. Eine erwähnenswerte Ausnahme ist die sogenannte Network Address Translation (NAT). Andererseits werden Schicht-2-Frames wie Ethernet immer umgeschrieben, wenn ein Router durchlaufen wird. Das Paket aus Abbildung 2-15 wurde außerdem erweitert, nachdem es auf Knoten A eingegangen ist, was die Änderung in der Adressierung auf der zweiten Schicht zeigt.

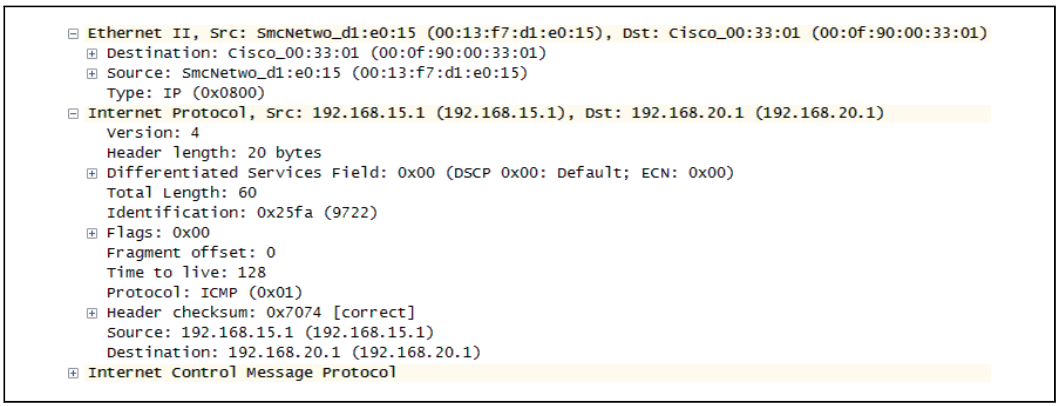
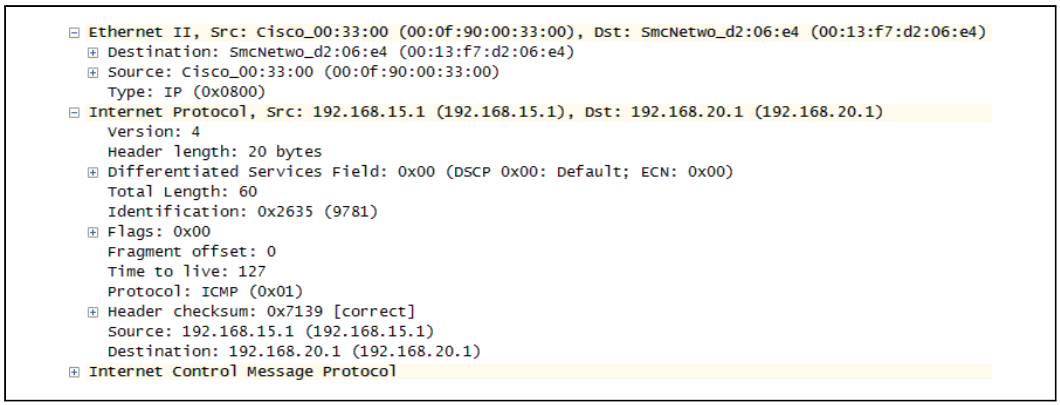


Abbildung 2-15 ▲
Im 2. Fall verwendete IP- und MAC-Adressen

Auf der anderen Seite (`192.168.20.0`) des Routers erfolgt die gleiche Art der Substitution. Leitet der Router den ICMP Echo Request an Knoten C weiter, entspricht die Quell-MAC-Adresse der des Routers und die des Ziels Knoten C, doch die IP-Adressen bleiben gleich. Denken Sie daran, dass der Router in dieser Topologie zwei Interfaces hat. Die MAC-Adressen der Schnittstellen werden ge-

nutzt und müssen unterschiedlich sein. Das ist in Abbildung 2-16 zu sehen. Beim Rückweg werden die MAC-Adressen vertauscht, d.h., Knoten C sendet auf dieser Seite etwas an das Router-Interface. Zurück im 192.168.15.0-Netzwerk läuft der Frame vom Router-Interface dieser Seite nach Knoten A.



▲ **Abbildung 2-16**
Im Netzwerk 192.168.20.0 verwendete Adressierung

Paket-Tracking

Der obige Abschnitt zum Thema Adressierung hat die Paket-Forwarding-Mechanismen und die auf beiden Seiten des Routers verwendeten Adressen erläutert. Dieser Abschnitt erläutert alle involvierten Pakete und setzt eine leere ARP-Tabelle voraus.

1. Fall: Das Ziel liegt im gleichen Netzwerk wie die Quelle

Unabhängig vom Ziel muss ein Knoten zuerst die lokale Host-Routingtabelle verarbeiten, um nach einem Treffer zu suchen. Da Knoten A (192.168.15.1) und Knoten B (192.168.15.2) im gleichen Netzwerk liegen, wird hier ein Treffer gefunden, bevor der oberste Eintrag in der Tabelle (das Standard-Gateway) erreicht ist. Der sendende Knoten entdeckt, dass die Interface- und Gateway-Adressen (die Spalten 3 und 4) identisch sind. Das heißt, der Zielknoten kann direkt kontaktiert werden. Da seine ARP-Tabelle leer ist, tauscht der sendende Knoten die folgenden Pakete aus:

▼ **Abbildung 2-17**
Kommunikation im Netzwerk 192.168.15.0

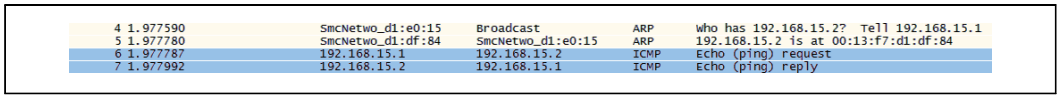


Abbildung 2-17 zeigt den ARP-Request von `192.168.15.1` und das darauf folgende ARP-Reply-Paket. Knoten A hat nun die MAC-Adresse der Ziel-IP gelernt und kann ein Ethernet-Frame erzeugen, in dem der ICMP Echo Request gekapselt wird. Sobald das ICMP Echo gesendet wurde, kommt eine Antwort von `192.168.15.2`.

2. Fall: Das Ziel liegt in einem anderen Netzwerk als die Quelle

Der zweite Fall ist etwas komplizierter als der erste, doch es gelten die gleichen Regeln. Die Kommunikation wird auf einer Seite des Routers initiiert, geht aber auf der anderen Seite weiter. Erneut muss Knoten A (`192.168.15.1`) seine Host-Routingtabelle verarbeiten. Diesmal entspricht der passende Eintrag dem Standard-Gateway. Knoten A bestimmt die Gateway-Adresse aus der dritten Spalte. Da seine ARP-Tabelle leer ist, tauscht der sendende Knoten folgende Pakete aus:

```
6 5.127439    SmcNetwo_d1:e0:15 Broadcast         ARP    Who has 192.168.15.254?  Tell 192.168.15.1
7 5.128503    Cisco_00:33:01    SmcNetwo_d1:e0:15 ARP    192.168.15.254 is at 00:0f:90:00:33:01
8 5.128510    192.168.15.1      192.168.20.1      ICMP   Echo (ping) request
```

Abbildung 2-18 ▲
Netzwerküberschreitende Kommunikation auf der Seite von 192.168.15.0

Knoten A sendet den ARP-Request für die MAC-Adresse des Routers und leitet, sobald er den ARP-Reply empfangen hat, den ICMP Echo Request zur Verarbeitung an den Router weiter. Auf der anderen Seite muss der Router nun Knoten C (`192.168.20.1`) finden und die Nachricht dorthin weiterleiten. Das ist in Abbildung 2-19 zu sehen.

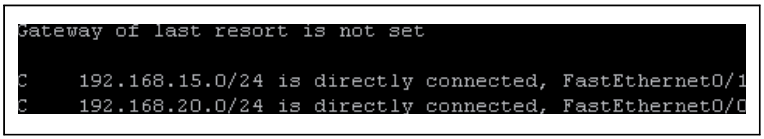
```
4 1.005457    Cisco_00:33:00 Broadcast      ARP    Who has 192.168.20.1?  Tell 192.168.20.254
5 1.005463    SmcNetwo_d2:06 Cisco_00:33:00 ARP    192.168.20.1 is at 00:13:f7:d2:06:e4
6 1.697588    192.168.15.1   192.168.20.1   ICMP   Echo (ping) request
7 1.697612    192.168.20.1   192.168.15.1   ICMP   Echo (ping) reply
```

Abbildung 2-19 ▲
Netzwerküberschreitende Kommunikation auf der Seite von 192.168.20.0

Der Router führt teilweise die gleichen Arbeiten durch wie der Host. Er bestimmt mittels ARP die MAC-Adresse von Host C und leitet, sobald er die MAC-Adresse kennt, den ursprünglichen ICMP Request weiter, indem er einen passenden Ethernet-Frame erzeugt. Knoten C empfängt den Request und muss ein ICMP Echo Reply an Knoten A zurückschicken. Der gesamte Prozess geht von vorne los,

aber diesmal in der entgegengesetzten Richtung. Der einzige Unterschied ist, dass einige ARP-Tabellen bereits gefüllt sind.

Hinweis: Der Vollständigkeit halber sei erwähnt, dass der Router einen zusätzlichen Schritt tun muss – er muss ebenfalls seine Routingtabelle verarbeiten. Die Tabelle ist in Abbildung 2-20 zu sehen.



◀ **Abbildung 2-20**
Routingtabelle des Routers

Lektüre

Dieses Kapitel ist wie Kapitel 1 nicht protokollspezifisch, auch wenn hier eine ganze Reihe von Protokollen vorgestellt wurden. Unabhängig vom Ziel besteht der erste Schritt immer darin, die Host-Routingtabelle zu verarbeiten. Nachdem das geschehen ist und der richtige Eintrag gefunden wurde, ist die Adressauflösung abgeschlossen. Das Verständnis der Host-Routingtabelle kann durch die Dokumentation des Betriebssystems verbessert werden. Eine hilfreiche Quelle ist das Microsoft Developer Network oder Technet (*http://technet.microsoft.com/en-us/*). Als leichte Lektüre habe ich einige Protokoll-RFCs aufgeführt, die bei diesem Prozess verwendet werden.

RFC 791: »Internet Protocol DARPA Internet Program Protocol Specification«
RFC 792: »Internet Control Message Protocol«
RFC 796: »Address Mapping«, J. Postel
RFC 826: »Ethernet Address Resolution Protocol«
RFC 894: »A Standard for the Transmission of IP Datagrams over Ethernet Networks«, C. Hornig
RFC 895: »A Standard for the Transmission of IP Datagrams over Experimental Ethernet Networks«, J. Postel
RFC 917: »Internet Subnets«
RFC 950: »Internet Standard Subnetting Procedure«
RFC 1256: »ICMP Router Discovery Messages«
RFC 1338: »Supernetting: an Address Assignment and Aggregation Strategy«
RFC 1519: »CIDR: an Address Assignment and Aggregation Strategy«

Zusammenfassung

Das Host-basierte Routing verlangt eine ganze Reihe eingebundener Prozesse. Host-Routingtabellen, Adressauflösung, Maskierung, Ethernet-Header und das Internet-Protokoll sind alle Teil dieser Geschichte. Die Folge der einzelnen Schritte hängt von der Lage des Ziels ab und kann die Konstruktion der Frames und die Verarbeitung beeinflussen. Wie alle Netzwerktabellen kann auch die Host-Routingtabelle manipuliert werden. Diese Prozesse und deren wechselseitige Beziehungen zu verstehen ist für eine gute Netzwerk-Administration von wesentlicher Bedeutung, da es bei der Fehlersuche, der Optimierung und der Sicherheit hilft.

Fragen

1. Was muss ein Knoten als Erstes tun, bevor er ein Paket senden kann?
2. Welche Meldung ist wahrscheinlich, wenn das Standard-Gateway fehlt?
3. Was ergibt die UND-Verknüpfung einer binären 1 mit einer binären 0?
4. Hosts suchen mit ARP immer nach Zielen in ihrem Netzwerk. Wahr oder falsch?
5. Kann ein Host ARP auch für ein Ziel nutzen, das nicht in seinem Netzwerk liegt?
6. Router modifizieren niemals den Ethernet-Frame, aber üblicherweise die IP-Header.
   a. WAHR
   b. FALSCH
7. Was sind bei einer Host-Routingtabelle die identifizierenden Eigenschaften der Gateway- und Interface-Spalten für einen Host im gleichen Netzwerk?
8. Nutzen Router das Address Resolution Protocol?
9. Multihome-Hosts sind Netzwerk-Knoten mit mehr als einem aktiven Interface und besitzen mehr Einträge in der Host-Routingtabelle.
   a. WAHR
   b. FALSCH

10. Wie viele Pakete werden im Idealfall bei einer leeren ARP-Tabelle generiert, wenn Knoten A den Befehl `ping -n 1 192.168.20.1` ausführt? Warum?

Antworten

1. Er muss die Host-Routingtabelle verarbeiten.
2. Zielhost nicht erreichbar (»Destination host unreachable«)
3. 0
4. Falsch. Wenn ein Eintrag in der ARP-Tabelle existiert, wird kein ARP-Request generiert.
5. Nein
6. FALSCH
7. Die Spalten enthalten die gleiche IP-Adresse.
8. Ja
9. WAHR
10. 8 – zwei ARP-Konversationen und zwei Folgen von ICMP Echo-Konversationen

Laborübungen

Übung 1: Aufbau der Topologie aus Abbildung 2-2

Material: Router und drei Hosts

1. Richten Sie die IP-Adressen ein. Wenn Sie den Router mit IP-Adressen versorgen, baut dies automatisch die Router-Routingtabelle auf.
2. Wenn Sie ein Linksys oder ein vergleichbares Gerät verwenden, müssen Sie sich die Optionen zur NAT-Deaktivierung ansehen. Standardmäßig verwenden Heim-Gateways NAT und Firewalls und maskieren oder unterdrücken den Netzwerk-Traffic.
3. Sobald Sie fertig sind, testen Sie Ihr Netzwerk, indem Sie jedes Gerät anpingen.

Übung 2: Host-Routingtabelle

Material: Knoten A

1. Untersuchen Sie die Host-Routingtabelle auf Knoten A mithilfe des MS Windows-Befehls `route print`.

2. Wählen Sie zwei oder drei Ziele aus. Bestimmen Sie per UND-Verknüpfung, welche Einträge der Routingtabelle den von Ihnen gewählten Zielen entsprechen.
3. Können Sie die zu verwendenden Gateways und Interfaces bestimmen?

Übung 3: ARP-Tabellen

Material: Router, Host-Befehlszeile und ARP-Tabellen

1. Ermitteln Sie den Inhalt der ARP-Tabellen mit dem Befehl `arp -a`.
2. Halten Sie die MAC- und IP-Adressen fest, falls welche vorhanden sind.
3. Löschen Sie die ARP-Tabellen. Der entsprechende Befehl variiert je nach Gerät, doch der Befehl `arp` ohne Argument gibt eine Hilfe aus. Stellen Sie sicher, dass die ARP-Tabelle des Routers ebenfalls gelöscht wird.
4. Warum ist dieses Experiment wichtig? Damit Sie den vollständigen Paketfluss und alle Prozesse sehen können, müssen alle Tabellen leer sein. Um Switch-SATs müssen Sie sich bei diesem Experiment keine Gedanken machen.

Übung 4: Traffic verfolgen

Material: Router, drei Hosts und Wireshark

1. Starten Sie das Paket-Capturing auf allen drei Knoten.
2. Pingen Sie zwischen Knoten A und den Knoten B und C.
3. Stellen Sie sicher, dass Sie ARP- und ICMP-Traffic festgehalten haben.
4. Nutzen Sie dieses Kapitel als Leitfaden, und schauen Sie, ob Sie dem Datenfluss zwischen den Knoten folgen können. Achten Sie auf die Timestamps, da die Dinge sehr schnell gehen – insbesondere dann, wenn Sie versuchen, den Traffic über Router hinweg zu verfolgen.
5. Gibt es Traffic, der nicht zu der von Ihnen generierten Kommunikation gehört? Warum gibt es ihn? Typische Beispiele sind Windows-Traffic, DNS und IPv6-Multicast.

Übung 5: Adressierung

Material: Router, drei Hosts und Wireshark

1. Sobald Sie den Traffic festgehalten haben, öffnen Sie die Pakete, und untersuchen Sie die verwendeten Adressen.
2. Erkennen Sie die Änderung in der Adressierung beim Wechsel des Routers von einer Seite auf die andere?
3. Welche Adressen haben sich geändert?
4. Ändern sich irgendwelche Adressen beim Pingen zwischen Knoten A und Knoten B?

KAPITEL 3

Spanning Tree und Rapid Spanning Tree

In diesem Kapitel:

- Warum sind Schleifen schlecht?
- Die Struktur von Spanning Tree-BPDUs
- Der Betrieb von Spanning Tree
- Spanning Tree-Nachrichten
- Cisco-Verbesserungen
- VLANs und Spanning Tree
- Rapid Spanning Tree Protocol
- Sicherheit
- Lektüre
- Zusammenfassung
- Fragen
- Antworten
- Laborübungen

Die Struktur und den Betrieb von Ethernet haben wir gut im Griff, weil das Basisprotokoll von einer Version zur nächsten konsistent bleibt und weil sich der Standard in nahezu jeder Topologie vorhersehbar verhält. Da viele Entscheidungen bei Ethernet – wie etwa die Netzwerkschnittstelle, die Signalisierung und der Gerätetyp – vorherbestimmt sind, könnte man glauben, dass der Ethernet-Einsatz einfach und geradlinig ist. Doch der saubere Betrieb eines Ethernet-Netzwerks hängt auch von der Einhaltung bestimmter Topologie-Regeln und anderen Protokollen (z.B. dem Address Resolution Protocol) ab. Ein einfaches Netzwerk entwickelt daher einige interessante und manchmal auch komplexe Charakteristika.

Dieses Kapitel behandelt das Spanning Tree Protocol (STP) und dessen schnellere Variante, das Rapid Spanning Tree Protocol (RSTP). Diese Protokolle führen einen fortwährenden Krieg gegen Schleifen in Ethernet-Netzwerken. Eine Schleife tritt in einem Ethernet-Netzwerk auf, wenn sich die Topologie mit sich selbst verbindet. Das ist ein Problem, weil Ethernet im Gegensatz zum Internet-Protokoll auf der 3. Schicht keinen entsprechenden Schutz integriert hat. Es kann daher nicht verhindern, dass Frames kontinuierlich zirkulieren. Tritt eine solche Schleife auf, kann die Konnektivität deutlich gestört oder sogar ganz unterbrochen werden.

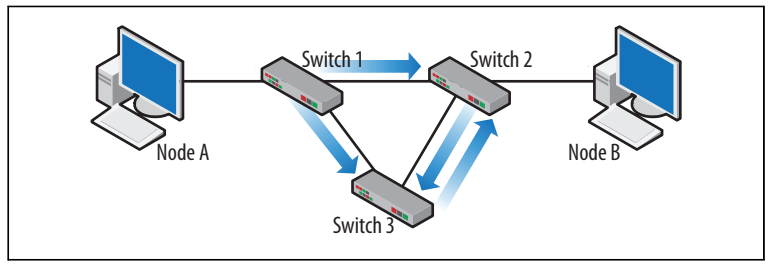
Das Spanning Tree Protocol ist standardmäßig aktiv und für Netzwerkadministratoren und Nutzer gleichermaßen unsichtbar. Doch dass es funktioniert und standardmäßig aktiv ist, heißt noch lange nicht, dass wir es ignorieren können. Manchmal ist Spanning Tree sehr ineffizient. »Standardmäßig aktiv« bedeutet auch, dass das Protokoll seinen Dienst hinter den Kulissen verrichtet. Spanning

Tree kann Aktionen durchgeführt haben, die den Administrator Probleme im Netzwerk vergessen lassen können. In diesem Kapitel wollen wir die Verwendung, den Betrieb und Sicherheitsaspekte von Spanning Tree behandeln. Das Spanning Tree Protocol ist in IEEE 802.1D standardisiert. Auch wenn die erste Version des Spanning Tree Protocol durch Rapid Spanning Tree abgelöst wurde, ist diese frühe Version dennoch häufig die Voreinstellung. Das heißt, ist immer noch wichtig, den früheren Standard zu verstehen. Heutzutage läuft Spanning Tree auf den meisten Bridges und Switches (mit Ausnahme einiger Wireless-Geräte).

Warum sind Schleifen schlecht?

Das Grundproblem besteht darin, dass Ethernet in der zweiten Schicht nicht die Möglichkeit hat, kontinuierlich zirkulierende Frames zu entfernen oder Schleifen zu verhindern. Im Gegensatz zu IP, das ein TTL-(Time to Live-)Feld besitzt, leiten Ethernet-Geräte wie Hubs oder Switches Frames einfach immer weiter, auch wenn eine Schleife vorliegt. Auf den ersten Blick mag das keine große Sache sein, doch wenn man bedenkt, dass ein einzelner Frame, der einen Switch passiert, mehrere Kopien erzeugen kann, wird die Bedeutung offensichtlich. Sehen wir uns eine kleine Topologie an. In Abbildung 3-1 sind drei Switches in einer Schleife verschaltet.

Abbildung 3-1 ▶
Switchingschleife



Wenn Knoten A mit Knoten B kommuniziert, ist der erste gesendete Frame ein (Broadcast-)ARP-Request. Das Standardverhalten des Switches besteht darin, diesen Frame an alle Ports zu senden, außer an den, der ihn gesendet hat. In diesem Fall sendet Switch 1 den Frame sowohl an Switch 2 als auch an Switch 3. Switch 2 und Switch 3 geben diesen Frame sofort gegenseitig an sich weiter. Und gleich darauf senden sie ihn direkt an Switch 1 zurück. Switch 1 besitzt nun zwei Kopien des ursprünglich von ihm gesendeten Frames, und um die Sache noch schlimmer zu machen, weiß er nicht, dass es sich um Kopien handelt. Also leitet Switch 1 diese

Kopien gleich wieder an Switch 2 und 3 weiter. Und so weiter, und so weiter ... Switches, die normalerweise Dutzende Frames pro Sekunde übertragen, können nun gezwungen sein, Hunderte oder gar Tausende Frames pro Sekunde zu übertragen (nur damit Sie eine Vorstellung davon haben, wie übel das werden kann). Die Auslastung des Backplanes kann dann in weniger als einer Minute von 10 % auf über 80 % ansteigen. Ruft man sich den Aufbau von Ethernet-Frames und das Weiterleitungsverhalten der Schicht-1- und -2-Geräte in Erinnerung, gibt es nichts,was dieses Problem lösen könnte. Da kommt das Spanning Tree Protocol ins Spiel.

Radia Perlman ist die Frau, der wir den Spanning Tree-Algorithmus verdanken. Man erzählt sich die Geschichte, dass sie das Problem während ihrer Arbeit bei der Digital Equipment Corp (DEC) erkannte und nach Hause ging, um darüber nachzudenken. Sie löste es an einem Samstag und hatte am Sonntag noch Zeit, ein Gedicht darüber zu schreiben. Wir sehen uns das Protokoll später an, aber hier ist schon mal das Gedicht.

Algorhyme
I think that I shall never see
A graph more lovely than a tree.
A tree whose crucial property
Is loop-free connectivity.
A tree which must be sure to span.
So packets can reach every LAN.
First the Root must be selected
By ID it is elected.
Least cost paths from Root are traced
In the tree these paths are placed.
A mesh is made by folks like me
Then bridges find a spanning tree.

Die Struktur von Spanning Tree-BPDUs

Spanning Tree verlangt von den Switches das Senden von Frames, die als Bridge Protocol Data Units (BPDUs) bezeichnet werden. Die in diesen BPDUs enthaltenen Informationen müssen von benachbarten Switches empfangen und verarbeitet werden. Die grundlegende Struktur ist in Abbildung 3-2 zu sehen.

Eine BPDU besteht aus drei Abschnitten: Details zum Protokoll, Felder für den Vergleichsalgorithmus und Timer-Werte. Jeder dieser Abschnitte wird nachfolgend genauer erläutert, doch für den Einstieg wurde dieser Frame in einem 802.3-Frame gekapselt. Management-Frames wie das Cisco Discovery Protocol nutzen häufig die 802.3-Kapselung, während Datenframes Ethernet Type II nutzen.

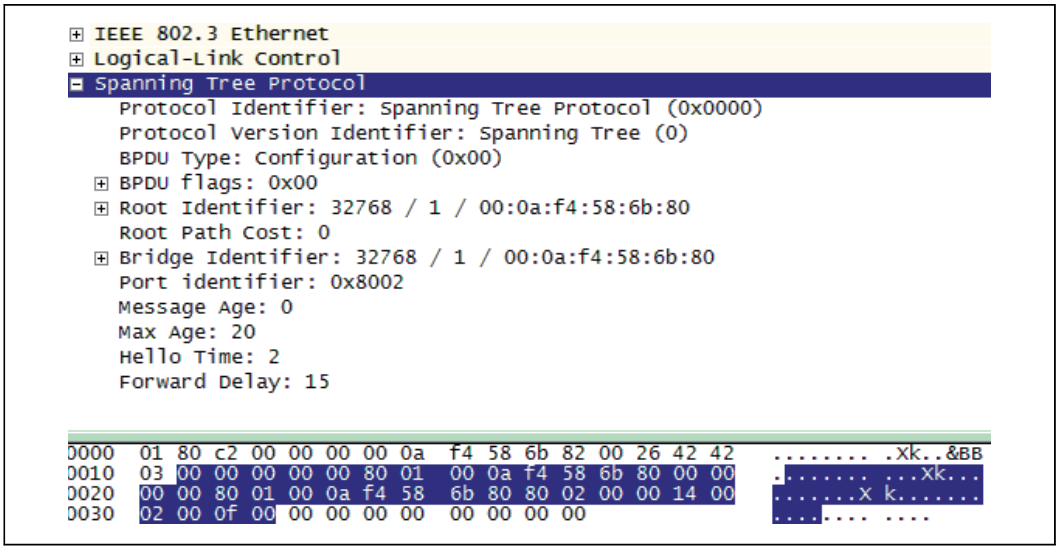


Abbildung 3-2 ▲
Bridge Protocol Data Unit

Der Vergleichsalgorithmus

Der ganze Sinn von Spanning Tree besteht darin, Schleifen zu verhindern, indem Ports im Netzwerk automatisch blockiert werden. Welche Ports blockiert werden müssen, wird durch einen Vergleichsalgorithmus bestimmt. Dieser Vergleichsalgorithmus verwendet bis zu vier Felder für den Vergleich: Root-Identifier, Root-Pfad-Kosten (root path cost), Bridge-Identifier (sendende(r) Bridge/Switch) und Port-Identifier (sendender Port). Aus Sicht von Spanning Tree sind kleinere Zahlen besser. Die Reihenfolge ist wichtig, und der Root-Identifier wird zuerst ermittelt. Abbildung 3-2 zeigt ein dekodiertes Paket zusammen mit seiner hexadezimalen Variante. Der Spanning Tree-Header ist hervorgehoben, um die entsprechenden Hexadezimalwerte zu zeigen. Wireshark schafft etwas Klarheit im Bezug auf den Inhalt der BPDU, d.h., es fügt einige Informationen hinzu, die im eigentlichen Frame nicht enthalten sind.

Die Informationen in diesen vier Feldern werden mit Informationen »verglichen«, die dem Switch bereits bekannt sind. Die Vergleiche werden verwendet, um Entscheidungen zur Steuerung in Schleifen-Topologien zu treffen. Spanning Tree erzwingt eine logische Topologie des Netzwerks, indem es Ports daran hindert, Datenframes zu senden. Das bedeutet, dass sich die physischen und logischen Topologien unterscheiden können.

Die Funktionen der vier Felder sind wie folgt:

Root-Identifier

Ein 8 Byte langes Feld, das eine Kombination aus der Priorität der Root-Bridge und deren MAC-Adresse darstellt. Ein typischer Wert für die Bridge-Priorität ist 32768 (8000 hexadezimal). Die VLAN-ID kann auf diesen Wert aufaddiert werden. Da alle Ports auf einem Cisco-Switch in VLAN 1 beginnen, ändert sich die Priorität zu 32768 + 1 (32769). Das entspricht hexadezimal 8001. In einer konvergierten oder stabilen Topologie haben alle BPDUs die gleiche Root-ID. In Abbildung 3-2 zeigt die dekodierte Ansicht die Root-ID 32768/1/000af4586b80. Schaut man sich den Hex-Bereich an, beginnt der Wert 8001000af4586b80 nach den ersten fünf Bytes. Der Unterschied entsteht durch die Verschmelzung von Priorität und VLAN-ID.

Root-Pfad-Kosten

Ein vier Byte langes Feld, das die Distanz zur Wurzel über die Anzahl und die Geschwindigkeit der Links beschreibt. In Abbildung 3-2 liegen die Pfadkosten bei 0, d.h., wir sehen uns eine BPDU an, die direkt von der Root-Bridge stammt. Die Werte für die Link-Geschwindigkeit sind:

10BaseT	100
100BaseT	19
1000Baset	4

Von der Root-Bridge ausgehende BPDUs haben unabhängig von der Link-Geschwindigkeit Pfadkosten von 0. Alle anderen BPDUs enthalten topologiebasierte Werte. Bei einem 100BaseT-Netzwerk würden BPDUs, die zwei Switches weiter liegen, Root-Pfad-Kosten von 38 aufweisen (siehe Abbildung 3-3).

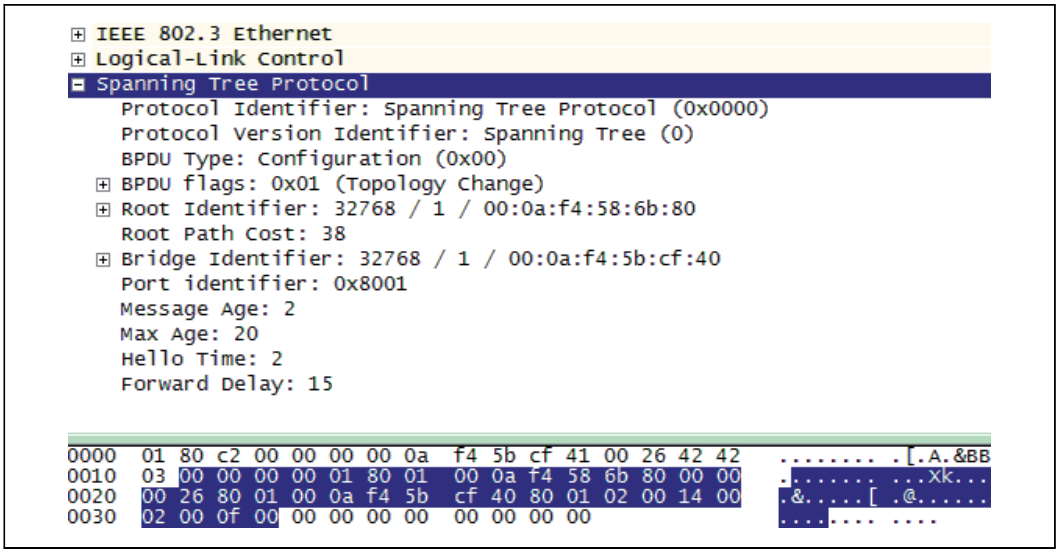


Abbildung 3-3 ▲
Erhöhte Pfadkosten

Bridge-Identifier

Dieses 8-Byte-Feld ist eine Kombination aus der MAC-Adresse der sendenden Bridge und ihrer Priorität. »Sendend« ist hier das Schlüsselwort, da sich der Wert auf den Switch bezieht, der die BPDU sendet. Auch hier lautet ein typischer Wert für die Bridge-Priorität 32768 (8000 hex) plus eventuelle VLANs. Der Switch, der die aktuelle BPDU sendet, trägt hier seine eigenen Werte ein. Abbildung 3-3 enthält eine BPDU aus dem gleichen Netzwerk wie die BPDU in Abbildung 3-2. Diese ist aber weiter von der Root-Bridge entfernt. In Abbildung 3-3 haben sich nicht nur die Pfadkosten erhöht, auch die Bridge-ID hat sich geändert. Das Root-ID-Feld bleibt gleich, da sich alle Switches in der Topologie auf diesen Wert verständigt haben. Die Bridge-ID lautet nun 32768/1/000af45bcf40 (8001000af45bcf40 hex), da der Frame nicht vom Root-Switch gesendet wurde. In Abbildung 3-2 sind Root-ID und Bridge-ID identisch – ein weiterer Hinweis darauf, dass die BPDU von der Wurzel stammt.

Port-Identifier

Der Port-Identifier ist das letzte Feld im Vergleichsalgorithmus. Diese zwei Bytes sind eine Kombination aus der Priorität des sendenden Ports und der Portnummer. Ein gängiger Wert für die Port-Priorität ist 128 (80 hex). In Abbildung 3-2 sehen wir den Wert 8002, d.h., die BPDU stammt von Port 2. Abbildung 3-3 enthält den Wert 8001, was bedeutet, dass die Switches zwar die gleiche Priorität haben, aber über verschiedene Ports gesendet

wurden und, aufgrund der unterschiedlichen Bridge-IDs, von verschiedenen Switches.

Die erste Aufgabe des Algorithmus besteht darin, die Root-Bridge zu bestimmen. Das ist eine einfache Prozedur, bei der die Kombination aus niedrigster Priorität und MAC-Adresse zur Root-Bridge erklärt wird. Beginnen alle Switches mit der gleichen Priorität (was üblich ist), wird der Switch mit der niedrigsten MAC-Adresse zur Root-Bridge. Es spielt keine Rolle, welcher Switch den Prozess anstößt, da die Switches BPDUs austauschen und sich die Spanning Tree-Topologie basierend auf den empfangenen Informationen ändern kann. Nachdem die Root-Brigde bestimmt wurde, werden designierte Bridges gewählt, Port-Rollen festgelegt und Ports blockiert, um Schleifen zu eliminieren. Die folgenden Abschnitte behandeln zuerst die Bausteine des Protokolls und gehen dann die operativen Aspekte durch, die diese Bausteine zusammenfügen.

Einige Definitionen

Innerhalb des Spanning Tree Protocol werden verschiedene Begriffe verwendet. Sie zu verstehen hilft bei den Topologie-Beispielen:

Root-Bridge
: Die Root-Bridge ist der Switch mit dem kleinsten numerischen Wert für Priorität und MAC-Adresse.

Designierte Bridge
: Wenn Traffic, der ein Segment des Netzwerks verlässt, sich zum Root-Switch bewegt, kann er einen weiteren Switch durchlaufen (von diesem weitergeleitet werden). Dieser Switch ist die designierte Bridge für dieses Segment.

Root-Ports und designierte Ports
: Sobald sich die Topologie stabilisiert hat, besitzen alle dem Root-Switch nachgeschalteten Switches Ports, die dem Root-Switch näher sind, und solche, die weiter weg sind. Die näher liegenden Ports werden Root-Ports genannt. Weiter entfernte Ports werden als designierte Ports bezeichnet. Man kann es auch so sehen, dass die Root-Ports in Richtung Root weisen und dass Traffic auf seinem Weg zum Root-Switch durch diese Ports läuft. Es gibt pro Switch nur einen Root-Port. Designierte Ports zeigen vom Root-Switch weg, und der Traffic läuft auf seinem Weg zum Root-Port in diese Ports hinein. Alle Ports des Root-Switchs sind designierte Ports. Root-Port und designierter Port werden als *Port-Rollen* bezeichnet.

Spanning Tree-Adressierung

Spanning Tree verwendet einen bestimmten Satz von Adressen. In Abbildung 3-4 sind die Ethernet- und LCC- (Logical Link Control-)Header zu sehen, die das verdeutlichen. Das ist eine andere Ansicht des Frames aus Abbildung 3-3.

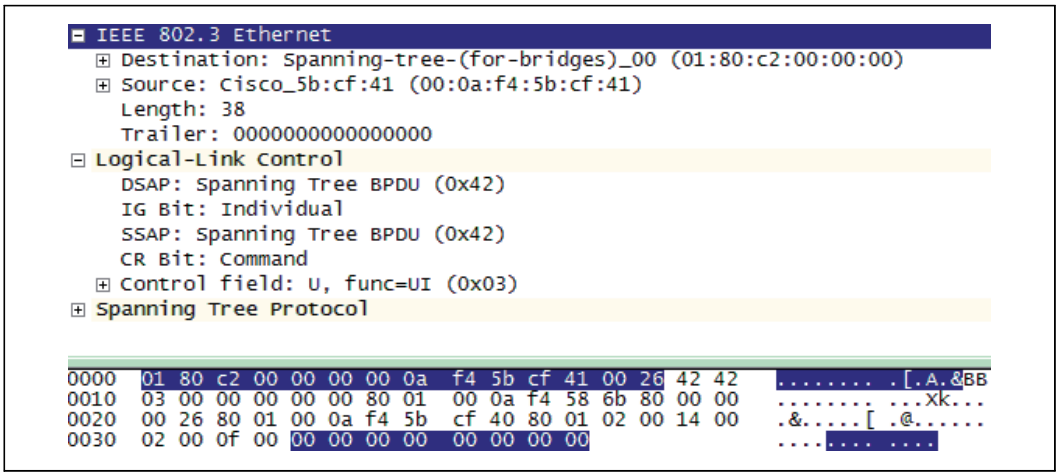


Abbildung 3-4 ▲
BPDU-Adressierung

Vergleichen Sie die Ethernet-Quelladresse aus Abbildung 3-4 (000af45bcf41) mit der Bridge-ID (8001000af45bcf40) aus Abbildung 3-3. Entfernt man die Bridge-Priorität, bleibt 000af45bcf40 übrig. Beachten Sie, dass die Ethernet-Quelladresse einfach um 1 inkrementiert wurde. Diese BPDU kam von Port 2 des Switches, weil die Port-ID 8001 lautet, was nun auch durch die MAC-Adresse verifiziert wird, da die Portnummer einfach zur MAC-Adresse des Ports hinzuaddiert wird. Würde diese BPDU von Port 10 stammen, würde die Port-ID in den Spanning Tree-Daten den Wert 800a enthalten und die Quell-MAC-Adresse im Ethernet-Frame wäre 000af45bcf4a. Das bedeutet also, dass der Switch und dessen einzelne Ports eindeutige MAC-Adressen besitzen.

Tipp

Gelegentlich muss ein Router in eine Bridge umgewandelt werden. In diesem Fall folgt die Adressierung nicht mehr dieser Konvention, und auch die Bridge-ID variiert.

Die Ethernet-Ziel-MAC-Adresse 0180c2000000 ist in der Brigde-Gruppen-Adresse definiert. Alle Switches und Bridges müssen diese Adresse kennen und an ihr »horchen« (Listening). Das ist auch der Grund, warum ein Cisco-Switch bei Spanning Tree-Operationen mit Switches anderer Hersteller kommunizieren kann. Diese Adres-

sen sind, zusammen mit dem LLC Destination Service Access Point (DSAP) und dem LLC Source Service Access Point (SSAP) in IEEE 802.1D für den Einsatz mit dem Spanning Tree Protocol festgelegt.

Port-Status

In einem laufenden Netzwerk achten Administratoren üblicherweise nicht auf den »Status« der Spanning Tree-Ports, weil verschiedene dieser Zustände nur Übergangszustände sind. Wenn man zu einem bestimmten Zeitpunkt schaut, befinden sich Traffic sendende/empfangende Ports bereits im »Forwarding«-Status. Alle Ports beginnen mit dem Status »blockiert«. Der Übergang zwischen den Zuständen wird durch einen Forwarding-Verzögerungstimer beeinflusst.

Blocked (blockierend)
: Ein Port mit diesem Status kann BPDUs empfangen, aber nicht senden. Datenframes werden nicht gesendet bzw. weitergeleitet. Ein solcher Port kann aber, je nachdem, welche STP-Informationen von benachbarten Bridges empfangen (oder nicht empfangen) wurden, mit dem Forwarding beginnen.

Listening (horchend)
: Das ist der erste Übergangszustand, in den eingetreten wird, wenn der Spanning Tree erkennt, dass der Port möglicherweise am Forwarding von Datenframes teilnehmen muss. Der Port empfängt und verarbeitet BPDUs, leitet aber keine Datenframes weiter. Bei diesem Status beginnen die Ports damit, BPDUs zu senden.

Learning (lernend)
: Dieser Status entspricht dem Horchen, nur dass der Port und der Switch nun die Topologie kennen und sich darauf vorbereiten, Datenframes weiterzuleiten. Der Port empfängt und verarbeitet auch weiterhin BPDUs.

Forwarding (weiterleitend)
: Der finale Status. Der Port leitet nun Datenframes weiter und verarbeitet gleichzeitig neue Informationen von eingehenden BPDUs.

Shutdown/Disabled (heruntergefahren/deaktiviert)
: Ein administrativ heruntergefahrener Port leitet weder Daten- noch Spanning Tree-(BPDU-)Frames weiter.

Spanning Tree-Timer

Ein großer Teil des Spanning Tree-Betriebs wird über eine Reihe von Timern kontrolliert. Die in einem Netzwerk verwendeten Werte sind am Ende der BPDU-Daten zu sehen (siehe Abbildung 3-3 und Abbildung 3-7).

Hello

Dieser Timer kontrolliert die Häufigkeit, mit der Konfigurations-BPDUs vom Root-Switch ausgegeben werden. Ein Standardwert sind 2 Sekunden. Das kann beim Capturing etwas lästig sein, da es so viele Hello-Timer gibt. Die Standard-BPDU in Abbildung 3-4 wird als »Hello«-Nachricht betrachtet.

Max Age (maximales Alter)

Switches halten nach, wie lange sie die aktuellen Informationen besitzen. Übersteigt das Alter dieser Informationen den Wert des Max-Age-Timers (20 Sekunden), muss der Vergleichsalgorithmus erneut ausgeführt werden. Der aktuelle Alters-Timer wird immer zurückgesetzt, sobald neue Informationen über eine BPDU empfangen werden. Ein Beispiel wäre ein benachbarter, BPDUs sendender Switch, der abgeklemmt wird. Der empfangende Switch erhält keine BPDUs mehr, und das Alter der aktuellen Informationen steigt. Schließlich muss der Empfänger einen neuen Pfad zur Wurzel finden.

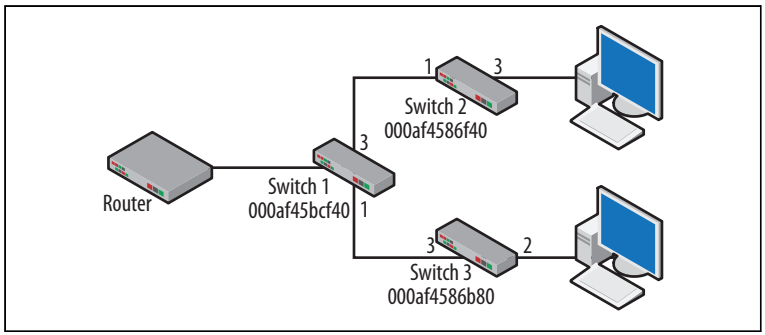
Forward Delay

Dieser Timer überwacht die Zeitspanne, die in den Übergangszuständen verbracht wird. Standard ist ein Limit von 15 Sekunden. Das erklärt auch die Verzögerung zwischen dem Anschließen eines Computers und dem Erscheinen eines Link-Lichts. Die Ports blockieren zuerst und warten 15 Sekunden auf BPDUs (Listening), und weitere 15 Sekunden vergehen bis zur Verarbeitung (Learning) und dem Forwarding. Die typische Verzögerung für das Link-Licht liegt also bei etwa 30 Sekunden.

Der Betrieb von Spanning Tree

Um den Betrieb von Spanning Tree besser verstehen zu können, wollen wir in einer kleinen Topologie ein Beispiel durchgehen. Die MAC-Adressen und BPDU-Werte werden die gleichen sein wie in den vorigen Paketen. Abbildung 3-5 zeigt die Topologie. Zu Beginn sind alle drei Switches ausgeschaltet. Der Router zeigt nur einen Weg aus dem Netzwerk heraus, ist in die Spanning Tree-Entschei-

dungen aber nicht involviert. Die Bridge-Prioritäten sind alle auf den Wert 32768 + VLAN oder 32769 (8001 hex) gesetzt (für VLAN 1), und die Port-Prioritäten sind mit 128 (80 hex) eingestellt. Der Vorgang beginnt mit dem Einschalten von Switch 1, dem dann weitere Switches folgen, was wir uns genauer ansehen wollen.



◀ **Abbildung 3-5**
Kleine Spanning Tree-Topologie

1. Schritt: Switch 1 wird eingeschaltet

Wird ein Cisco-Switch gebootet, leuchten alle Port-Link-Lichter gelblich, was bedeutet, dass sie momentan keinen Traffic weiterleiten. Schließt man einen Computer an einen Port an, gibt es außerdem eine kurze Verzögerung, bevor das Link-Licht grün wird. Das liegt daran, dass die Switch-Ports anfänglich üblicherweise »blockiert« sind. Der Switch muss zuerst etwas über die Netzwerktopologie lernen, bevor er damit beginnen kann, Daten weiterzuleiten. Auf diese Weise wird der Aufbau einer direkten Schleife verhindert.

Sobald Switch 1 potenziellen BDPU-Traffic empfangen und verarbeitet hat, kann er damit beginnen, Daten weiterzuleiten. Während dieser Zeit wechseln die Ports zwischen den verschiedenen Zuständen hin und her. Nutzt man bei einem Cisco-Switch den Befehl zum Debugging von Spanning Tree-Events, erhält man die folgende Ausgabe:

▼ **Abbildung 3-6**
Port-Zustände

```
Switch#
1d02h: %LINK-3-UPDOWN: Interface FastEthernet0/3, changed state to up
1d02h: set portid: VLAN0001 Fa0/3: new port id 8003
1d02h: STP: VLAN0001 Fa0/3 -> listening
1d02h: %LINEPROTO-5-UPDOWN: Line protocol on Interface FastEthernet0/3, changed state to up
1d02h: STP: VLAN0001 Fa0/3 -> learning
1d02h: STP: VLAN0001 Fa0/3 -> forwarding
Switch#
```

Während Port FastEthernet 0/3 (F0/3) hochkommt, durchläuft er die Zustände *blocking→listening→learning→forwarding*, wobei während der Listening- und Learning-Zustände eine Verzögerung von 15 Sekunden eintritt.

Von den Ports 1 und 3 ausgehend, würden wir die BPDUs aus Abbildung 3-7 sehen. Beachten Sie, dass die Ethernet-Quell-MAC-Adresse und die Port-IDs in den BPDU-Daten dem Port entsprechen, der die BPDU sendet. Da nur Switch 1 läuft, ist er Root-Switch und sendender Switch gleichermaßen. Das spiegelt sich in den BDPU-Root-ID- und Bridge-ID-Feldern wider. Die Pfadkosten für beide BPDUs sind 0.

Abbildung 3-7 ▼
BPDUs für Schritt 1

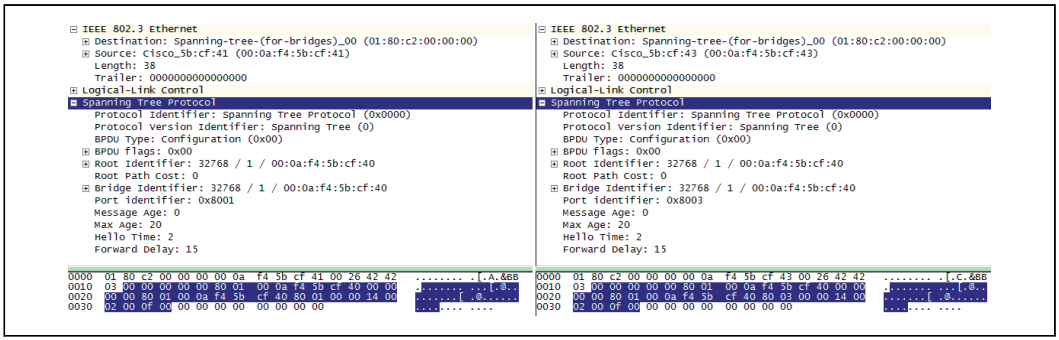


Abbildung 3-8 ▼
Spanning Tree für Schritt 1

Das Ergebnis dieses Schrittes kann man sich mit dem Befehl `show spanning-tree` ansehen.

```
Switch#
Switch#show spanning-tree

VLAN0001
  Spanning tree enabled protocol ieee
  Root ID     Priority    32769
              Address     000a.f45b.cf40
              This bridge is the root
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    32769  (priority 32768 sys-id-ext 1)
              Address     000a.f45b.cf40
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
              Aging Time 300

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Fa0/1            Desg FWD 19        128.1    Shr
Fa0/3            Desg FWD 19        128.3    Shr

Switch#
```

In Abbildung 3-8 sind die Root-ID und die Bridge-ID gleich. Die Bridge-Priorität ist mit 32769 angegeben. Port F0/1 und F0/3 leiten Daten mit einem Kostenfaktor von 19 (also mit 100 Mbps) und einer Port-Priorität von 128 weiter. Die Timer-Werte sind ebenfalls aufgeführt.

2. Schritt: Switch 2 wird eingeschaltet

Bei stabilen Zuständen fließen BPDUs immer von der Root-Bridge weg. Das liegt einfach daran, dass es keine Notwendigkeit gibt, flussaufwärts liegende Switches über die Netzwerkbedingungen zu informieren, da diese die Ausgangsquelle sind. Das gilt so lange, bis sich etwas im Netzwerk ändert. In diesem Fall ist die MAC-Adresse von Switch 2 kleiner als die von Switch 1. Das bedeutet, dass bei gleichen Bridge-Prioritäten (und sie sind gleich), Switch 2 zum neuen Root-Switch wird. Die Ports an Switch 2 blockieren, verarbeiten aber die von Switch 1 kommenden BPDUs. Switch 2 bemerkt, dass der im Root-ID-Feld enthaltene Wert unter seiner eigenen Bridge-ID liegt, und antwortet Switch 1 mit einer BPDU, die einen Coup d'Etat andeutet. Sieht man sich die Debugging-Ausgabe auf Switch 1 an, kann man seiner Reaktion folgen.

```
1d03h: STP: VLAN0001 heard root 32769-000a.f458.6b80 on Fa0/1
1d03h:     supersedes 32769-000a.f45b.cf40
1d03h: STP: VLAN0001 new root is 32769, 000a.f458.6b80 on port Fa0/1, cost 19
1d03h: STP: VLAN0001 sent Topology Change Notice on Fa0/1
```

▲ **Abbildung 3-9**
Die Debugging-Ausgabe für Switch 1 zeigt einen neuen Root-Switch.

Ein Capture der BPDUs zwischen Switch 1 und Switch 2 zeigt, dass die BPDUs nun von Switch 2 kommen statt von Switch 1. Das wichtige Detail ist hier, dass sich die Root- und Bride-IDs geändert haben. Erinnern Sie sich daran, dass diese identisch sind und dass diese BPDU daher vom Root-Switch stammt. Das wird durch die Root-Pfadkosten von 0 untermauert. Vergleichen Sie Abbildung 3-10 mit Abbildung 3-7.

```
⊞ IEEE 802.3 Ethernet
⊞ Logical-Link Control
⊟ Spanning Tree Protocol
    Protocol Identifier: Spanning Tree Protocol (0x0000)
    Protocol Version Identifier: Spanning Tree (0)
    BPDU Type: Configuration (0x00)
  ⊞ BPDU flags: 0x00
  ⊞ Root Identifier: 32768 / 1 / 00:0a:f4:58:6b:80
    Root Path Cost: 0
  ⊞ Bridge Identifier: 32768 / 1 / 00:0a:f4:58:6b:80
    Port identifier: 0x8003
    Message Age: 0
    Max Age: 20
    Hello Time: 2
    Forward Delay: 15

0000  01 80 c2 00 00 00 00 0a  f4 58 6b 83 00 26 42 42   ........ .Xk..&BB
0010  03 00 00 00 00 00 80 01  00 0a f4 58 6b 80 00 00   ........ ...Xk...
0020  00 00 80 01 00 0a f4 58  6b 80 80 03 00 00 14 00   .......X k.......
0030  02 00 0f 00 00 00 00 00  00 00 00 00               ....---- ----
```

Abbildung 3-10 ▲
BPDU von dem neuen Root-Switch

Das Ergebnis sieht man auch in den BPDUs, die von Switch 1 an die andere Seite der Topologie übertragen werden. In Abbildung 3-11 wurde die BPDU vom alten Root-Switch (Bridge-ID-Feld) gesendet, doch Switch 2 wird als neuer Root-Switch im Root-ID-Feld (32768/1/000af4586b80) angegeben.

Abbildung 3-11 ▼
Die BPDU vom alten Root-Switch gibt den neuen Root-Switch bekannt.

```
⊞ IEEE 802.3 Ethernet
⊞ Logical-Link Control
⊟ Spanning Tree Protocol
    Protocol Identifier: Spanning Tree Protocol (0x0000)
    Protocol Version Identifier: Spanning Tree (0)
    BPDU Type: Configuration (0x00)
  ⊞ BPDU flags: 0x00
  ⊞ Root Identifier: 32768 / 1 / 00:0a:f4:58:6b:80
    Root Path Cost: 19
  ⊞ Bridge Identifier: 32768 / 1 / 00:0a:f4:5b:cf:40
    Port identifier: 0x8003
    Message Age: 1
    Max Age: 20
    Hello Time: 2
    Forward Delay: 15

0000  01 80 c2 00 00 00 00 0a  f4 5b cf 43 00 26 42 42   ........ .[.C.&BB
0010  03 00 00 00 00 00 80 01  00 0a f4 58 6b 80 00 00   ........ ...Xk...
0020  00 13 80 01 00 0a f4 5b  cf 40 80 03 01 00 14 00   .......[ .@......
0030  02 00 0f 00 00 00 00 00  00 00 00 00               ....---- ----
```

Die Pfadkosten haben sich auf 19 erhöht, da der Traffic nun durch den 100-Mbps-Switch 1 laufen muss, um den Root-Switch zu erreichen. Schließlich zeigt die Zusammenfassung auf Switch 1 die Änderungen in der Topologie in anderer Form an.

```
Switch#
Switch#sh spanning-tree

VLAN0001
  Spanning tree enabled protocol ieee
  Root ID     Priority    32769
              Address     000a.f458.6b80
              Cost        19
              Port        1 (FastEthernet0/1)
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    32769  (priority 32768 sys-id-ext 1)
              Address     000a.f45b.cf40
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
              Aging Time 300

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------
Fa0/1            Root FWD 19        128.1    Shr
Fa0/3            Desg FWD 19        128.3    Shr

Switch#
```

▲ **Abbildung 3-12**
Switch 1 zeigt Spanning Tree mit neuem Root-Switch.

Abbildung 3-9 zeigt das Debugging-Ergebnis von Switch 1, der eine BPDU von Switch 2 empfängt. Dieses Ereignis ändert auch die Ausgabe des `show spanning-tree`-Befehls, wie man in Abbildung 3-12 sehen kann. Wie bei den meisten Netzwerk-Prozessen gewähren die Untersuchung von Paket-Captures und die Ausgabe der Netzwerk-Geräte einen Einblick in den Betrieb des Protokolls.

3. Schritt: Switch 3 wird eingeschaltet

Durch Hinzufügen von Switch 3 ist unsere Topologie vollständig, wenn auch noch nicht konvergiert. Basierend auf der MAC-Adresse und der Priorität erkennen alle drei Switches Switch 2 als Root an. BPDUs fließen von Switch 2 weg und geben die Topologie-Informationen so, wie sie gerade kommen, bekannt. Switch 3 ändert nicht sehr viel an der Topologie, außer dass er sie erweitert. Am weitesten Punkt von Switch 2 (der obere Computer in Abbildung 3-5) würde ein Caputre nun Pfadkosten zum Root-Switch von 38 anzeigen. Dabei ist Switch 3 der sendende Switch. Das ist in Abbildung 3-13 zu sehen.

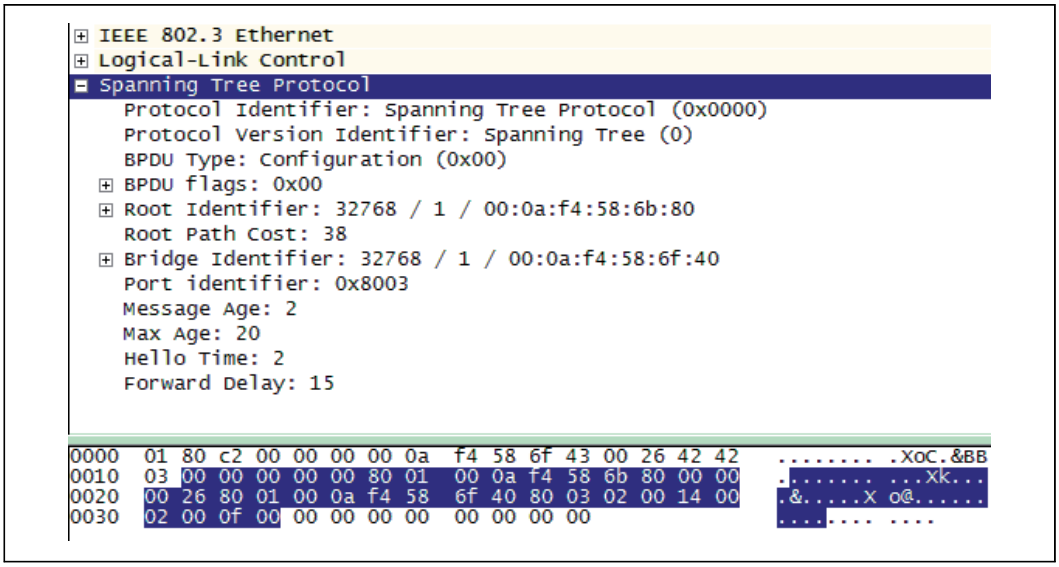


Abbildung 3-13 ▲
Von Switch 3 gesendete BPDU

Beachten Sie, dass sich die Bridge-ID geändert hat, nicht aber die Root-ID. Die Pfadkosten haben sich auf 38 erhöht. Die Port-ID ist 8003.

Nachdem man ein Verständnis für die grundlegende Struktur und den Betrieb von Spanning Tree entwickelt hat, erkennt man, dass es bei dieser kleinen Topologie nicht benötigt wird, da es keine Schleifen gibt. Doch was passiert, wenn jemand Switch 2 mit Switch 3 verbindet, sei es nun versehentlich (was ständig passiert) oder um etwas Redundanz zu schaffen? Dann geht der Spaß erst richtig los. Weiter geht es mit dem 4. Schritt.

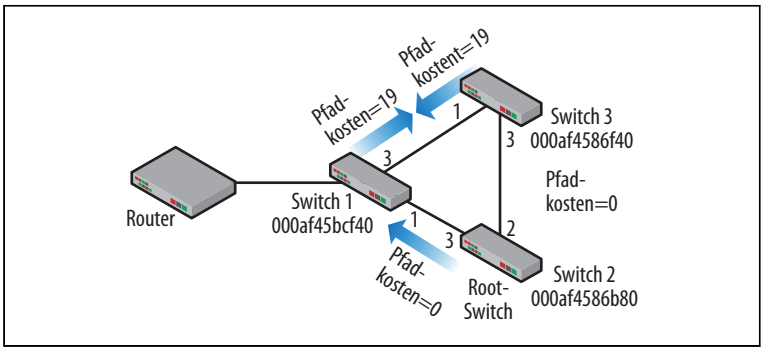
4. Schritt: Eine Schleife aufbauen

Wird eine physische Schleife aufgebaut, indem man die Switches 2 und 3 miteinander verbindet, reagiert Spanning Tree mit dem Blockieren eines der Ports der Topologie. Sind mehrere Schleifen vorhanden, werden weitere Ports blockiert, bis die Schleifen eliminiert sind. Zu Beginn sind die physischen und logischen Topologien identisch. Die Entscheidung, welcher Port blockiert wird, basiert vollständig auf den vom Vergleichsalgorithmus verwendeten Informationen. Doch zuerst müssen wir eine andere Frage klären: Wie werden Schleifen erkannt? Während des normalen Betriebs fließen die BPDUs vom Root-Switch weg. Anders ausgedrückt: Ein Switch sollte Informationen über den Root-Switch nur aus einer Richtung erhalten. »Hört« ein Switch etwas über den Root-Switch via die BPDUs mehrerer Ports, ist eine Schleife aufgetreten.

Unabhängig von der Spanning Tree-Version reagieren Switches sehr schnell, um die Schleife zu eliminieren. Während die Switches die BPDU-Informationen vergleichen, ist der Switch mit den niedrigsten Werten der Verlierer und muss einen Port blockieren. Abbildung 3-14 zeigt die neue Topologie mit der Schleife. Wir wissen, dass Switch 2 basierend auf den Prioritäten und MAC-Adressen zum Root-Switch wird. Daher haben alle BPDUs in diesem Netzwerk den gleichen Wert im ersten Feld des Vergleichsalgorithmus.

Root ID – 32768/1/000af4586b80

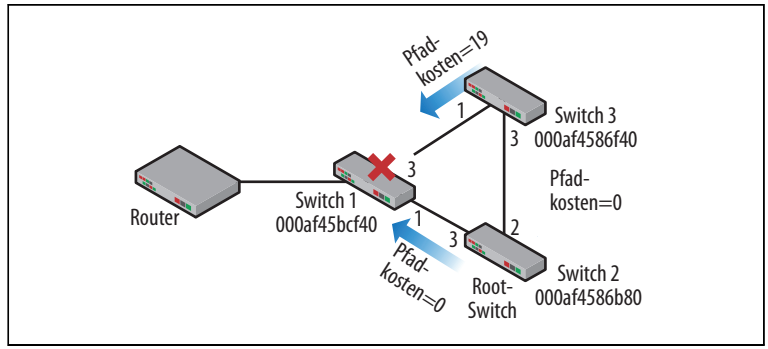
Als Nächstes sind die Pfadkosten zum Root-Switch zu berücksichtigen. Der Root-Switch sendet zu Pfadkosten von 0 an den Ports 2 und 3. Natürlich sind das die kleinstmöglichen Werte für die Pfadkosten, und Root wird nicht aufgefordert, einen Port zu sperren. Die Switches 1 und 3 stehen in Bezug auf Root-ID und Pfadkosten schlechter da. Es liegt also an Switch 1 und 3 zu entscheiden, welcher der Downstream-Ports zu blockieren ist. Dazu senden sie sich gegenseitig BPDUs.



◀ Abbildung 3-14
Topologie mit Schleife

Auf dem Link zwischen Switch 1 und Switch 3 werden die BPDUs mit der gleichen Root-ID und, wie in Abbildung 3-14 zu sehen ist, mit den gleichen Pfadkosten gesendet. In diesem Fall reduziert sich die Entscheidung, welcher Port blockiert wird, auf die Bridge-ID. Bei gleichen Bridge-Prioritäten »verliert« die höhere MAC-Adresse, und Switch 1 muss einen Port blockieren. Während sich die physische Topologie weiterhin wie in Abbildung 3-14 darstellt, verhält sich die logische Topologie nun wie in Abbildung 3-15. Port 3 von Switch 1 wurde hier blockiert und die Schleife eliminiert.

Abbildung 3-15 ▶
Aufgelöste Topologie

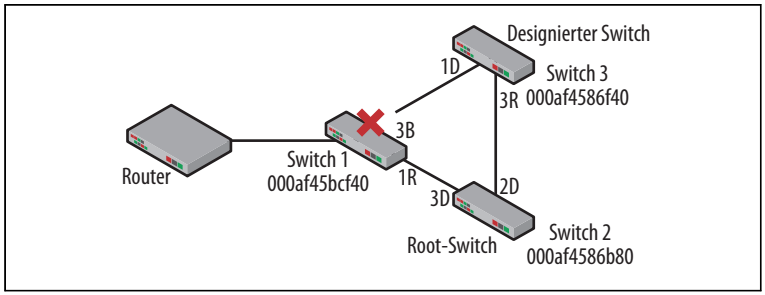


Die Ausgabe des `show spanning tree`-Befehls auf Switch 1 (siehe Abbildung 3-16) macht die Topologie-Änderungen deutlich. Denken Sie daran, dass das Kabel physisch immer noch vorhanden ist.

```
Switch#
2d02h: STP: VLAN0001 sent Topology Change Notice on Fa0/1
2d02h: STP: VLAN0001 Fa0/3 -> blocking
Switch#
Switch#sh span

VLAN0001
  Spanning tree enabled protocol ieee
  Root ID     Priority    32769
              Address     000a.f458.6b80
              Cost        19
              Port        1 (FastEthernet0/1)
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    32769  (priority 32768 sys-id-ext 1)
              Address     000a.f45b.cf40
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
              Aging Time 15

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------
Fa0/1            Root FWD 19        128.1    Shr
Fa0/3            Altn BLK 19        128.3    Shr

Switch#
```

Abbildung 3-16 ▲
Switch 1 zeigt Spanning Tree mit blockiertem Port.

Oben in Abbildung 3-16 wird das Ereignis festgehalten, und der finale Status wird unten mit der Interface-Liste ausgegeben. Root-Ports, designierte Ports, Root-Bridges und designierte Bridges wurden im vorigen Abschnitt definiert. Sobald sich die Topologie stabilisiert hat, können wir die Rollen jedes Netzwerk-Gerätes und jedes Ports klar erkennen. In Abbildung 3-17 stehen die Buchstaben B, R und D für blockierte, Root- und designierte Ports.



◀ **Abbildung 3-17**
Switch- und Port-Rollen

Diese Rollen kann man auch in der Ausgabe des `show spanning tree`-Befehls in Abbildung 3-12 und Abbildung 3-16 sehen.

Spanning Tree-Nachrichten

Um zu dieser neuen stabilen Konfiguration zu gelangen, müssen eine ganze Reihe von Informationen über BPDUs ausgetauscht werden. Das Spanning Tree Protocol kennt einen kleinen Satz von Nachrichten, darunter das bereits vorgestellte »hello«. Doch durch den Aufbau einer Schleife und die Notwendigkeit, einen Port zu blockieren, wird es Zeit, sich über die anderen Arten von Nachrichten zu unterhalten. Der ein Byte große BPDU-Typ wird basierend auf der verwendeten Nachricht gesetzt. Das ein Byte große BPDU-Flags-Feld gibt an, welche Operation gerade durchgeführt wird. Die Werte des stabilen Zustands sind in Abbildung 3-18 zu sehen. Dieser Frame wurde festgehalten, kurz bevor die Schleife aufgebaut wurde.

```
⊟ Spanning Tree Protocol
    Protocol Identifier: Spanning Tree Protocol (0x0000)
    Protocol Version Identifier: Spanning Tree (0)
    BPDU Type: Configuration (0x00)
  ⊟ BPDU flags: 0x00
      0... .... = Topology Change Acknowledgment: No
      .... ...0 = Topology Change: No
  ⊞ Root Identifier: 32768 / 1 / 00:0a:f4:58:6b:80
    Root Path Cost: 38
  ⊞ Bridge Identifier: 32768 / 1 / 00:0a:f4:58:6f:40
    Port identifier: 0x8003
    Message Age: 2
    Max Age: 20
    Hello Time: 2
    Forward Delay: 15
```

◀ **Abbildung 3-18**
BPDU-Typ und -Flags

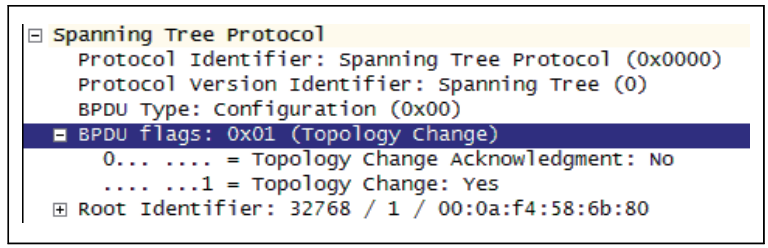
Configuration (Konfiguration)

Dies ist der Standard-Nachrichtentyp. Das BPDU-Typfeld ist auf 00 gesetzt. Die Nachricht enthält alle in diesem Kapitel diskutierten Informationen. Abbildung 3-18 ist ein Beispiel für eine Konfigurations-BPDU mit auf 00 gesetztem Flags-Feld.

Topology Change (Topologieänderung)

Diese BPDU zeigt an, dass eine Rekonfiguration der Topologie erfolgt. Als die Schleife aufgebaut wurde, haben die Switches 1 und 3 BPDUs ausgetauscht. Switch 3 hat erkannt, dass sein Pfad zum Root-Switch besser ist, und hat dementsprechend eine Topologieänderung (Topology Change oder kurz TC) initiiert. Eine ganze Reihe von Ereignissen können eine Topologieänderung in einer Spanning Tree-Topologie auslösen: der Ablauf des Max-Age-Timers, das Hinzufügen oder Entfernen eines Switches, das Hinzufügen oder Entfernen eines Links sowie der Empfang neuer Informationen über eine BPDU. In unserem Schleifen-Beispiel wurde die Änderung angestoßen. Switch 3 erkannte, dass es einen zweiten Weg zum Root-Switch gibt. Die Änderung im Flags-Feld ist in Abbildung 3-19 zu sehen.

Abbildung 3-19 ▶
Flags bei Topologieänderung



Topology Change Notification (Benachrichtigung über eine Topologieänderung)

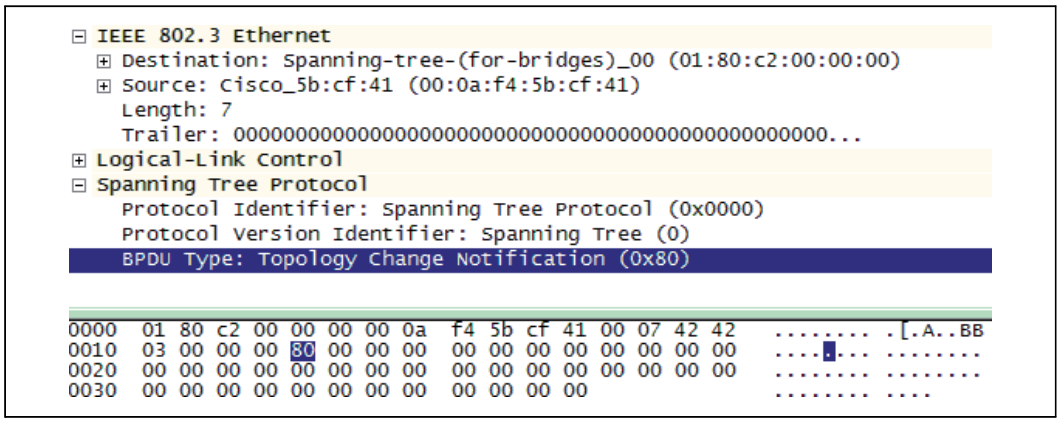
Beim Empfang einer BPDU von Switch 3 erkennt nun Switch 1, dass sich die Topologie ändert. BPDUs fließen in die gleiche Richtung, doch es gibt eine Schleife. Switch 1 sendet nun eine Topology Change Notification (TCN) an den Root-Switch zurück, was in Abbildung 3-20 zu sehen ist.

Die TCN-BPDU enthält keinerlei Konfigurationsinformationen. Ein genauer Blick auf die Hex-Darstellung der BPDU zeigt, dass der Vorspann (aufgefüllt mit Nullen) viel größer ist, um der minimalen Ethernet-Framegröße zu entsprechen.

Der Topologieänderungsprozess hält lange genug an, um den Ports innerhalb der Topologie den Wechsel in den richtigen Zustand (Forwarding oder Blockiert) zu ermöglichen. Entspre-

chend dem Standard läuft der Forwarding-Delay-Timer zweimal durch. Damit sind wir bei etwa 30 Sekunden. Sie können die Zahl der Konfigurations-BPDUs während dieser Zeitspanne tatsächlich messen und landen üblicherweise bei 15 oder 16.

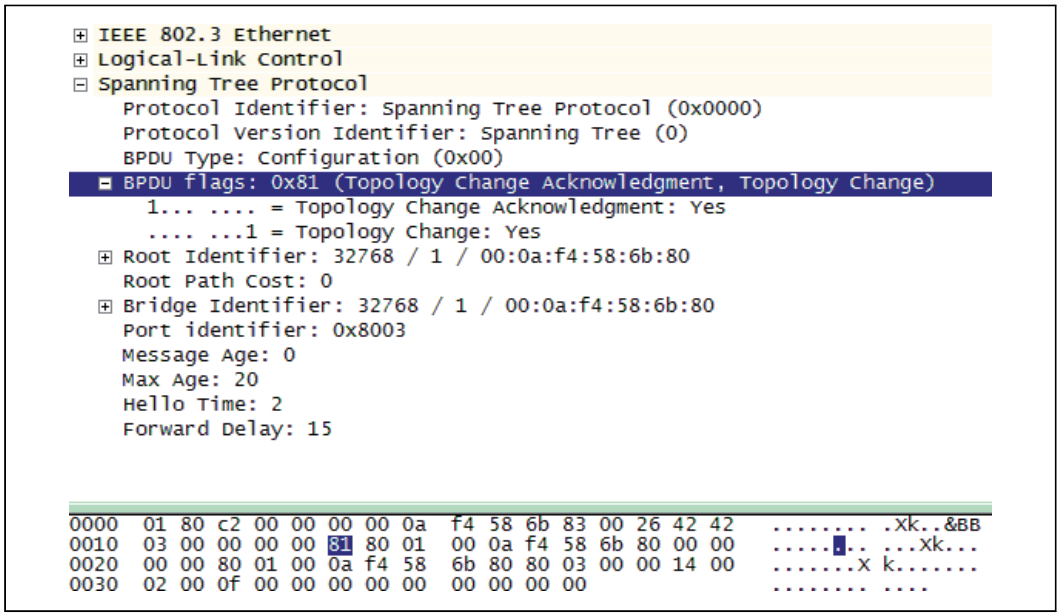
▼ **Abbildung 3-20**
Topology Change Notification



Topology Change Notification Acknowledgement (Bestätigung der TCN)

Wird eine TCN gesendet (in diesem Fall von 000af45bcf41 auf Switch 1), gibt der empfangende Switch eine Antwort in Form einer TCN ACK-Nachricht zurück (siehe Abbildung 3-21). Diese Nachricht gibt ihren Zweck über das Flags-Feld bekannt und liefert die aktuellsten Informationen zur Topologie zurück.

▼ **Abbildung 3-21**
TCN-Bestätigung

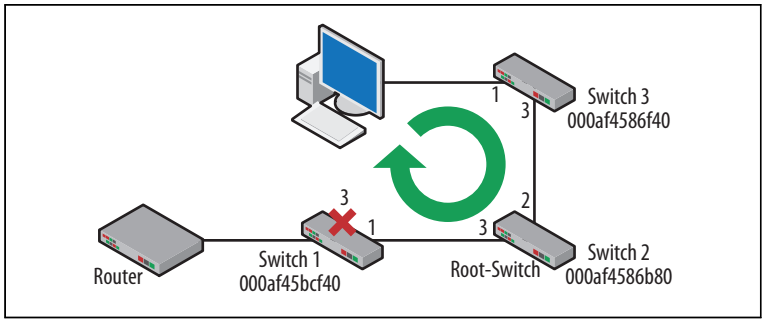


Eine letzte Anmerkung zum Flags-Feld ist die Tatsache, dass es zwar ein Byte lang ist, aber nur einige wenige Nachrichtentypen kennt. Wir behalten diese Information im Hinterkopf und kommen später in diesem Kapitel noch darauf zurück.

Probleme mit Spanning Tree

Spanning Tree ist bei der Eliminierung von Schleifen sehr gut. Es dauert in unserer Beispiel-Topologie weniger als fünf Sekunden, um das Problem zu beheben. Wenn allerdings Informationen verloren gehen oder bessere Pfade auftauchen, kann Spanning Tree furchtbar langsam sein. Wird beispielsweise die Schleife aus der Topologie entfernt, indem das Kabel zwischen Switch 1 und 2 gezogen wird (was den Zustand in Abbildung 3-5 wiederherstellt), wird Spanning Tree Port 4 an Switch 1 nicht sofort wieder »entsperren«. Stattdessen müssen wir warten, dass die von Switch 3 empfangenen Informationen veralten. Switch 3 würde nicht länger BPDUs vom Root-Switch empfangen. Der Max-Age-Timer läuft dann irgendwann ab, und die Topologieänderung würde beginnen. Doch wie lange dauert das? Der Wert des Max-Age-Timers liegt bei 20 Sekunden, und danach wird die TCN gesendet. Der Forwarding-Delay-Timer liegt für die Listening- und Learning-Zustände bei jeweils 15 Sekunden. Port 3 an Switch 1 bleibt also noch für 50 Sekunden blockiert, nachdem die Schleife bereits aufgelöst wurde. Jeder mit diesem Switch verbundene Knoten wäre für diese Zeitspanne isoliert. Wenn sich die Größe des Netzwerks oder dessen Komplexität erhöht, steigt auch die Zeitspanne an. Diese Verzögerung macht den Original-Spanning Tree für Redundanz-Lösungen ungeeignet.

Ein anderes Problem und einer der Gründe dafür, warum es gut ist, das Protokoll zu verstehen, ist, dass »automatische« Spanning Tree-Topologien in Bezug auf das Forwarding nicht optimal sind. Nehmen wir an, ein Host ist mit Switch 3 verbunden und versucht, eine Webseite zu öffnen. Basierend auf der ursprünglichen Blocking-Lösung (Post 3 an Switch 1), muss der Traffic über die gesamte Topologie laufen (siehe Abbildung 3-22).



◀ **Abbildung 3-22**
Suboptimales Forwarding

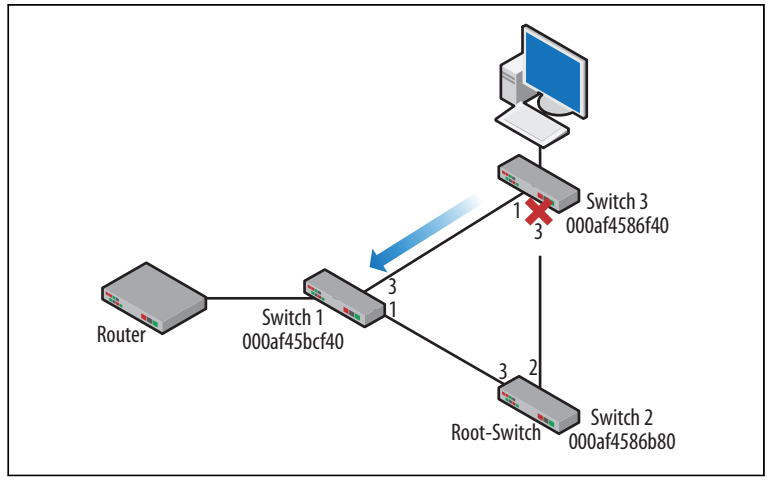
In diesem Fall hat Spanning Tree die Schleife eliminiert, doch Probleme bei der Verarbeitung von Traffic geschaffen. Um die Dinge für alle Knoten zu verbessern, könnte man die Priorität von Switch 1 so ändern, dass er zum Root-Switch wird. Erinnern Sie sich daran zurück, dass die Bridge-ID eine Kombination aus Bridge/Switch-Priorität, gefolgt von der MAC-Adresse ist. In Abbildung 3-23 wurde die Priorität auf 4096 + 1 für das VLAN (1001 hex) reduziert, wodurch Switch 1 zum neuen Root-Switch wird.

Die Topologie löst sich nun, wie in Abbildung 3-24 zu sehen ist, auf. Der Weg aus dem Netzwerk läuft jetzt für Knoten an den Switches 2 und 3 direkt über Switch 1. Port 3 an Switch 1 wurde entsperrt, und Port 3 an Switch 3 wurde blockiert.

▼ **Abbildung 3-23**
BPDU mit Prioritätsänderung

```
⊞ IEEE 802.3 Ethernet
⊞ Logical-Link Control
⊟ Spanning Tree Protocol
    Protocol Identifier: Spanning Tree Protocol (0x0000)
    Protocol Version Identifier: Spanning Tree (0)
    BPDU Type: Configuration (0x00)
  ⊞ BPDU flags: 0x01 (Topology Change)
  ⊞ Root Identifier: 4096 / 1 / 00:0a:f4:5b:cf:40
    Root Path Cost: 0
  ⊞ Bridge Identifier: 4096 / 1 / 00:0a:f4:5b:cf:40
    Port identifier: 0x8003
    Message Age: 0
    Max Age: 20
    Hello Time: 2
    Forward Delay: 15

0000  01 80 c2 00 00 00 00 0a  f4 5b cf 43 00 26 42 42   ........ .[.C.&BB
0010  03 00 00 00 00 01 10 01  00 0a f4 5b cf 40 00 00   ........ ...[.@..
0020  00 00 10 01 00 0a f4 5b  cf 40 80 03 00 00 14 00   .......[ .@......
0030  02 00 0f 00 00 00 00 00  00 00 00 00               ........ ....
```

Abbildung 3-24 ▶
Topologie mit neuem Root-Switch, basierend auf der Priorität

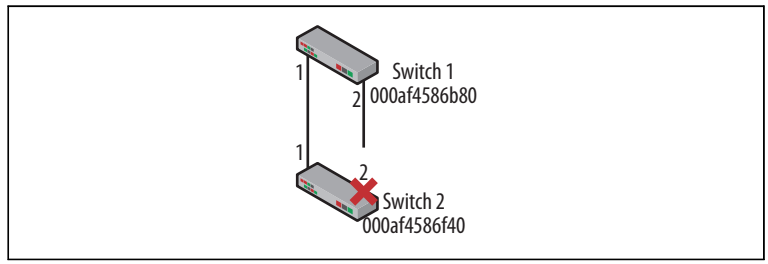


Switch zu Switch: Ein Sonderfall

Unabhängig von der Topologie beginnt der gesamte Prozess damit, dass ein Root-Switch basierend auf Priorität und MAC-Adresse gewählt wird. Danach werden die Switches mit den besten Pfadkosten zum Root-Switch zu den designierten Bridges für den Traffic, der aus einem bestimmten Netzwerksegment herausfließt. Im Fall eines Pfadkosten-Patts, wie wir es vorhin gesehen haben, »verliert« der Nicht-Root-Switch mit der kleinsten Bridge-ID (Priorität und MAC-Adresse) und muss einen Port blockieren. Doch wann wird das Port-ID-Feld verwendet?

In Abbildung 3-25 sind die beiden Switches in einer Schleife direkt miteinander verbunden. Wieder schreitet Spanning Tree ein und blockiert einen der Ports. Um zu bestimmen, welcher Port betroffen ist, muss der Vergleichsalgorithmus verwendet werden:

Abbildung 3-25 ▶
Sonderfall



1. Wähle einen Root-Switch. Unter der Annahme, dass die Prioritäten gleich sind (32768 + 1), ist die MAC-Adresse der entscheidende Faktor. Switch 1 wird zum Root-Switch.
2. Pfadkosten. BPDUs fließen vom Root-Switch weg, und da beide BPDUs den Root-Switch verlassen, liegen die beiden Pfadkosten bei 0.
3. Bridge-IDs vergleichen. Die Bridge-ID ist die ID der sendenden Bridge. In diesem Fall haben beide BPDUs den gleichen Wert (8001:000af4586b80), da beide vom Root-Switch kommen.
4. Port-IDs vergleichen. Das ist unsere letzte Möglichkeit, um die Schleife zu verhindern. Alle anderen Felder haben in den jeweiligen BDPUs die gleichen Werte. Wenn wir nun die BPDUs vergleichen, die die Ports 1 und 2 verlassen, sehen wir endlich einen Unterschied. Von Port 1 hat die Port-ID den Wert 8001 (Priorität 128 und Portnummer 1), während die von Port 2 an Switch 1 kommende BPDU die Port-ID 8002 (Priorität 128 und Portnummer 2) aufweist.

Aufgrund der von Switch 1 empfangenen Information entscheidet Switch 2 nun, seinen eigenen Port 2 zu blockieren, wodurch die Schleife aufgelöst wird. Es ist wichtig zu bemerken, dass die in den BPDUs enthaltenen Informationen nichts mit Switch 2 zu tun haben.

Cisco-Verbesserungen

Das Rapid Spanning Tree Protocol (RSTP) ist seit über einem Jahrzehnt Teil des Standards. Allerdings wurde es von den Geräten der einzelnen Hersteller nicht immer unterstützt. Selbst wenn Ihre Geräte RSTP unterstützen, sind sie mit älteren Bridges und Switches möglicherweise nicht kompatibel. Cisco hat eine ganze Reihe von Verbesserungen an STP vorgenommen, in dem Bemühen, die Konvergenzgeschwindigkeit und das Port-Forwarding zu verbessern.

Portfast

Spanning Tree ist für Bridges und Switches gedacht. Hosts kümmern sich nicht besonders um die Netzwerk-Topologie. Wenn also ein Host mit dem Netzwerk verbunden ist, ist es nicht nötig, ihn auf Übergänge in den Port-Zuständen warten zu lassen. Tatsächlich können Geräte wie VoIP-Telefone darunter sogar leiden. Das liegt daran, dass das Telefon in einer frühen Phase des Verbindungsprozesses eine ganze Reihe von Transaktionen abzuschließen versucht.

Der Befehl `spanning-tree portfast` informiert den Port darüber, dass er die Listening- und Learning-Zustände nicht durchlaufen muss. Das ist sehr praktisch, wenn Sie in einer dynamischen Umgebung Fehler suchen oder testen. Allerdings darf das nur für Endknoten verwendet werden. Wenn Sie ein mit Portfast konfiguriertes Interface versehentlich mit einem anderen Switch verbinden, kann es zu Schleifen kommen. Auf die potenziellen Gefahren wird von Cisco hingewiesen, wenn Sie den Befehl ausführen.

Abbildung 3-26 ▼
Portfast-Warnung

```
Switch(config-if)#spanning-tree portfast
%Warning: portfast should only be enabled on ports connected to a single
 host. Connecting hubs, concentrators, switches, bridges, etc... to this
 interface  when portfast is enabled, can cause temporary bridging loops.
 Use with CAUTION

%Portfast has been configured on FastEthernet0/11 but will only
 have effect when the interface is in a non-trunking mode.
Switch(config-if)#
```

Uplinkfast

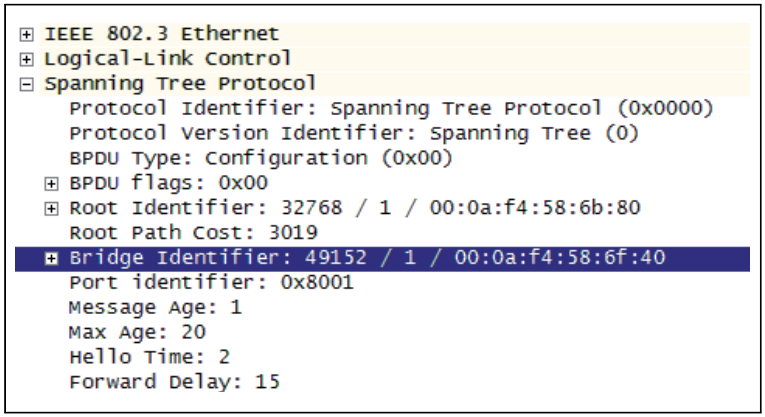
Uplinkfast soll die Konvergenzgeschwindigkeit verbessern, wenn es alternative Pfade zum Root-Switch gibt. Eine Schwäche von Spanning Tree sind lange Konvergenzverzögerungen aufgrund der Timer, auch wenn Ausweichpfade existieren. Selbst bei der kleinen vorhin diskutierten Topologie lag die Konvergenzzeit bis zur Behebung der Schleife bei 50 Sekunden.

Bei dieser Topologie führt man auf den Switches 1 und 3 den Befehl `spanning-tree uplinkfast` aus, wodurch sie mit ihren Nachbarn zu Mitgliedern einer Uplink-Gruppe werden. Es gibt eine Reihe von Änderungen in der BPDU; Bridge-Priorität und Pfadkosten werden erhöht. Das stellt sicher, dass sie *nicht* zum Root-Switch werden, da sie nun bestimmte Rollen übernehmen. Abbildung 3-27 zeigt die BPDU-Änderungen.

Ein weiterer Blick auf die Ausgabe von `show spanning-tree` offenbart, dass der Switch eine völlig andere Sicht der Topologie hat. Die in der BPDU enthaltene Änderung ist vorhanden, doch zusätzlich wird eine Alternative aufgeführt. Um genau zu sein, wurde der blockierte Port bereits vorher als Alternative geführt (siehe Abbildung 3-16), aber mit längeren Verzögerungen betrieben.

Mit konfiguriertem Uplinkfast wechselt der Switch sofort zum sekundären Pfad über Switch 3, wenn der Link zum Root-Switch verloren geht. Zusätzlich beginnt Port 3 an Switch 1 mit dem For-

warding. Das dauert nur etwa eine Sekunde und nicht wie vorhin fast eine Minute.



◀ Abbildung 3-27
Uplinkfast-BPDU

▼ Abbildung 3-28
Uplinkfast-Ausgabe

```
Switch#sh spanning-tree uplinkfast
UplinkFast is enabled

Station update rate set to 150 packets/sec.

UplinkFast statistics
-----------------------
Number of transitions via uplinkFast (all VLANs)            : 0
Number of proxy multicast addresses transmitted (all VLANs) : 0

Name                 Interface List
-------------------- ------------------------------------
VLAN0001             Fa0/1(fwd), Fa0/3
Switch#
Switch#sh span

VLAN0001
  Spanning tree enabled protocol ieee
  Root ID     Priority    32769
              Address     000a.f458.6b80
              Cost        3019
              Port        1 (FastEthernet0/1)
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    49153  (priority 49152 sys-id-ext 1)
              Address     000a.f45b.cf40
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
              Aging Time 300
  Uplinkfast enabled

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Fa0/1            Root FWD 3019      128.1    Shr
Fa0/3            Altn BLK 3019      128.3    Shr
```

Backbonefast

Das Ziel von Uplinkfast ist eine schnellere Konvergenzzeit für einen Switch, der die Verbindung zum Root-Switch verloren hat. Doch was ist mit einem beliebigen Switch innerhalb der Topologie, der die Verbindung zum Root-Switch verliert? Normalerweise heißt es »jeder Switch für sich«, und die Max-Age-Timer müssen zuerst ablaufen, bevor etwas getan werden kann. In der Topologie aus Abbildung 3-17 ist Switch 2 beispielsweise der Root-Switch, und Switch 1 blockiert Port 3, um eine Schleife zu eliminieren. Uplinkfast hat beim Verlust der Verbindung zwischen Switch 1 und 3 geholfen. Wenn nun aber der Link zwischen Switch 2 (Root) und Switch 3 verloren geht, kann Switch 1 irgendwie helfen?

Verliert Switch 3 die Verbindung zum Root-Switch, kann er sich mittels Backbonefast als Root anbieten, indem er eine entsprechende BPDU an Switch 1 sendet. Switch 1 ist aber immer noch mit dem ursprünglichen Root-Switch verbunden und betrachtet daher die BPDU als schlechter. Switch 1 kann nun einen speziellen Frame, den sogenannten Root Link Query (RLQ) senden, um herauszufinden, ob der Root-Switch immer noch vorhanden ist. Ist der Root-Switch immer noch aktiv, kann Switch 1 den blockierten Port 3 direkt in den Listening-Status bringen, ohne auf das Ablaufen des Max-Age-Timers warten zu müssen. Das spart etwa 20 Sekunden Konvergenzzeit gegenüber dem normalen Spanning Tree.

Abbildung 3-29 ▼
Root Link Query

```
⊞ IEEE 802.3 Ethernet
⊟ Logical-Link Control
    DSAP: SNAP (0xaa)
    IG Bit: Individual
    SSAP: SNAP (0xaa)
    CR Bit: Command
  ⊟ Control field: U, func=UI (0x03)
      000. 00.. = Command: Unnumbered Information (0x00)
      .... ..11 = Frame type: Unnumbered frame (0x03)
    Organization Code: Cisco (0x00000c)
    PID: Unknown (0x0108)
⊟ Data (35 bytes)
    Data: 00000000018001000af4586b80000000138001000af45bcf...
    [Length: 35]

0000  01 80 c2 00 00 00 00 0a  f4 5b cf 41 00 2b aa aa   ........ .[.A.+..
0010  03 00 00 0c 01 08 00 00  00 00 01 80 01 00 0a f4   ........ ........
0020  58 6b 80 00 00 00 13 80  01 00 0a f4 5b cf 40 80   Xk...... ....[.@.
0030  01 02 00 14 00 02 00 0f  00 00 00 00               ........ ....
```

Wireshark dekodiert den Root Link Query nicht vollständig. Denken Sie daran, dass das hochgradig proprietär ist und üblicherweise nicht angewandt wird. Dennoch zeigt uns eine Untersuchung der Datenfelder die MAC-Adressen der involvierten Switches. Noch aufschlussreicher ist die Kommunikation zwischen den Switches. Dieser Frame wird von einem Nicht-Root-Switch generiert, wenn er die schwächere BPDU empfängt. Dem folgt eine RLQ-Antwort vom Root-Switch. Sobald der Pfad zum Root-Switch feststeht, wechselt der blockierte Port an Switch 1 sofort in den Listening-Zustand.

VLANs und Spanning Tree

An anderer Stelle in diesem Kapitel gab es Anzeichen dafür, dass STP durch VLANs beeinflusst wird. Die Bridge-Priorität berücksichtigt die VLAN-ID, indem es ihren Wert zu 32768 hinzuaddiert. Laut Kapitel 4 sind VLANs separate IP-Netzwerke, die als separate Schicht-2-Broadcast-Domains bestehen. Es zeigt sich, dass jedes VLAN eine eigene Instanz von Spanning Tree laufen haben kann. Das bedeutet, dass jedes VLAN bei Bedarf eine andere logische Topologie aufweisen kann als die anderen, auf dem gleichen Switch laufenden VLANs. Im Cisco-Sprachgebrauch wird das als »Per VLAN Spanning Tree« oder PVST bezeichnet.

Sehen wir uns die Topologie in Abbildung 3-30 an. Oberflächlich betrachtet, ist das die gleiche Topologie wie vorhin, nur dass nun VLANs auf den Switches laufen und dass die Switches über Trunks miteinander verbunden sind. Ich habe auch einige Server für die VLANs 4 und 5 eingefügt. Standardmäßig würde Spanning Tree eine Topologie aufbauen, die (selbst mit den VLANs) genau dem entspricht, was wir bisher gesehen haben. Das liegt daran, dass Spanning Tree seine Entscheidungen selbst mit VLANs auf die gleiche Weise trifft. Obwohl es separate VLANs gibt, sind die logischen Topologien identisch, auch wenn sie unabgängig berechnet werden.

Das Problem besteht darin, dass selbst bei dieser kleinen Topologie die Nutzer von VLAN 4 an Switch 3 das ganze Netzwerk durchlaufen müssen, um auf die Server zugreifen zu können. Für die Benutzer von VLAN 5 gilt das nicht, weil die Server in der Mitte des Netzwerks liegen. Um das effizienter zu gestalten, können wir Switch 1 zum Root-Switch für VLAN 4 machen.

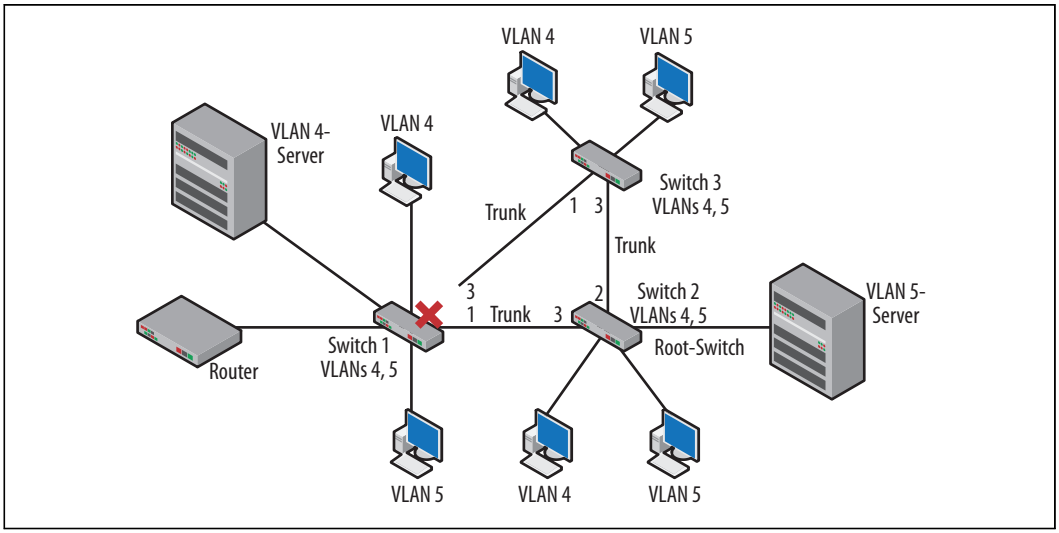
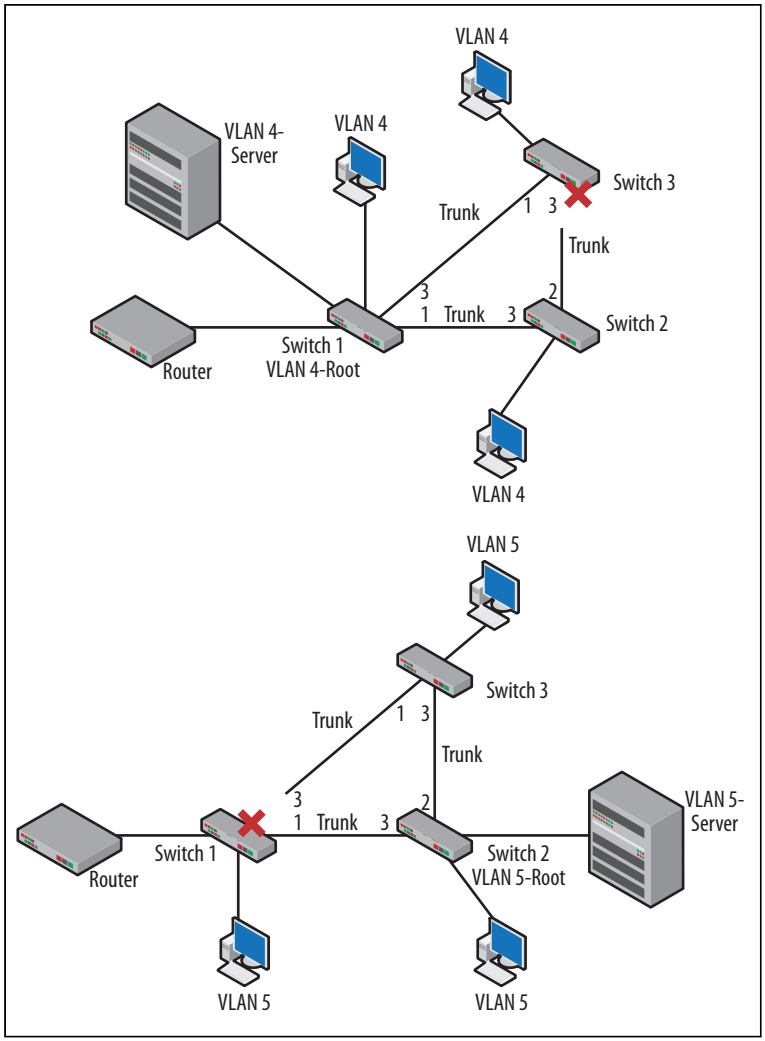


Abbildung 3-30 ▲
Spanning Tree-Topologie mit VLANs

Wie schon zuvor, nehmen wir eine kleine Änderung basierend auf Prioritäten vor. In Abbildung 3-24 ist das Ergebnis der Änderung der Switch-Prioritäten in der BPDU zu sehen. Nicht so offensichtlich ist, dass die Prioritätsänderung tatsächlich für das VLAN erfolgt. Genauer gesagt war das eine Änderung für VLAN 1, weil in diesem Fall alle Ports in VLAN 1 liegen. Der verwendete Befehl war `spanning tree vlan 1 priority 4096`. Für dieses Beispiel ändern wir den Befehl zu `spanning tree vlan 4 priority 4096`. Beachten Sie, dass wir VLAN 5 nicht anpacken, da die Root für VLAN 5 genau da ist, wo sie sein soll. Im Endergebnis ändert sich die physische Topologie nicht, doch es gibt nun *zwei* logische Topologien, wie Abbildung 3-31 zeigt.



◀ **Abbildung 3-31**
Neue Spanning Tree-Topologien, basierend auf VLAN

Auf der linken Seite sehen wir, dass die Topologie so modifiziert wurde, dass Port 3 an Switch 3 blockiert wird. Diese Modifikation bringt die Ressourcen näher an die Nutzer von VLAN 4 heran. Die rechte Seite der Topologie bleibt unverändert. Schaut man sich die `show spanning tree`-Ausgabe von vorhin (Abbildung 3-16 und Abbildung 3-28) an, dann erkennt man die VLAN-ID als Teil der Ausgabe. Für die Topologien in Abbildung 3-31 würde die Ausgabe des `show spanning tree`-Befehls die beiden Konfigurationen durch separate Ausgaben für die VLANs 4 und 5 anzeigen. Es ist wichtig zu erkennen, dass die Paket-Forwarding-Entscheidungen nicht nur

auf den MAC-Adressen basieren, sondern auch auf den VLAN-IDs der Trunks. Diese Details werden in Kapitel 4 behandelt.

Ein weiterer wichtiger Vorteil mehrerer Spanning Tree-Instanzen besteht darin, dass für einen Teil des logischen Traffics Ports verwendet werden können, die ansonsten blockiert sein könnten. Das verbessert den Durchsatz und die Performance.

Rapid Spanning Tree Protocol

Das Spanning Tree Protocol gemäß IEEE 802.1D ist bei der Eliminierung von Schleifen sehr effektiv, in anderen Situationen, etwa bei der Suche nach neuen Pfaden, ist es aber sehr langsam. Die meisten Organisationen verlassen sich bei Redundanz, Load-Balancing oder der Ausfallsicherung nicht auf Spanning Tree oder andere Schicht-2-Lösungen. Switches werden häufig durch Router ersetzt, wenn diese Fähigkeiten gefragt sind. Cisco hat Verbesserungen wie Portfast, Uplinkfast und Backbonefast eingeführt, um der langsamen Konvergenz bzw. den Port-Zustandswechseln entgegenzuwirken.

Das Rapid Spanning Tree Protocol (RSTP), das in IEEE 802.1w standardisiert ist, erhöht die Performance des ursprünglichen Spanning Trees und fügt viele der Funktionen hinzu, die man aus den Cisco-Verbesserungen kennt. Darüber hinaus ist es schnell genug, um eine verlässliche Komponente robuster und hochverfügbarer Netzwerke zu sein.

Zu den signifikanten Änderungen zählen:

- Switches verwerfen schnell die alten Daten, wenn neue Daten empfangen werden.
- Mehrere neue Port-Rollen wurden definiert.
- Die Port-Zustände wurden ebenfalls geändert.

Beim Spanning Tree nach 802.1D gibt es innerhalb der BDPU keinen Hinweis auf die Rolle des Ports, sobald er mit dem Forwarding begonnen hat. Aus den BPDUs, die wir uns vorhin angesehen haben, wissen wir, dass die BPDU-Typ- und -Flags-Felder aus zwei Bytes bestehen, dass gleichzeitig aber nur wenige Werte oder Typen verwendet werden. RSTP nutzt diese Felder, um zusätzliche Informationen zu den Ports (und damit zur Topologie) zu übermitteln.

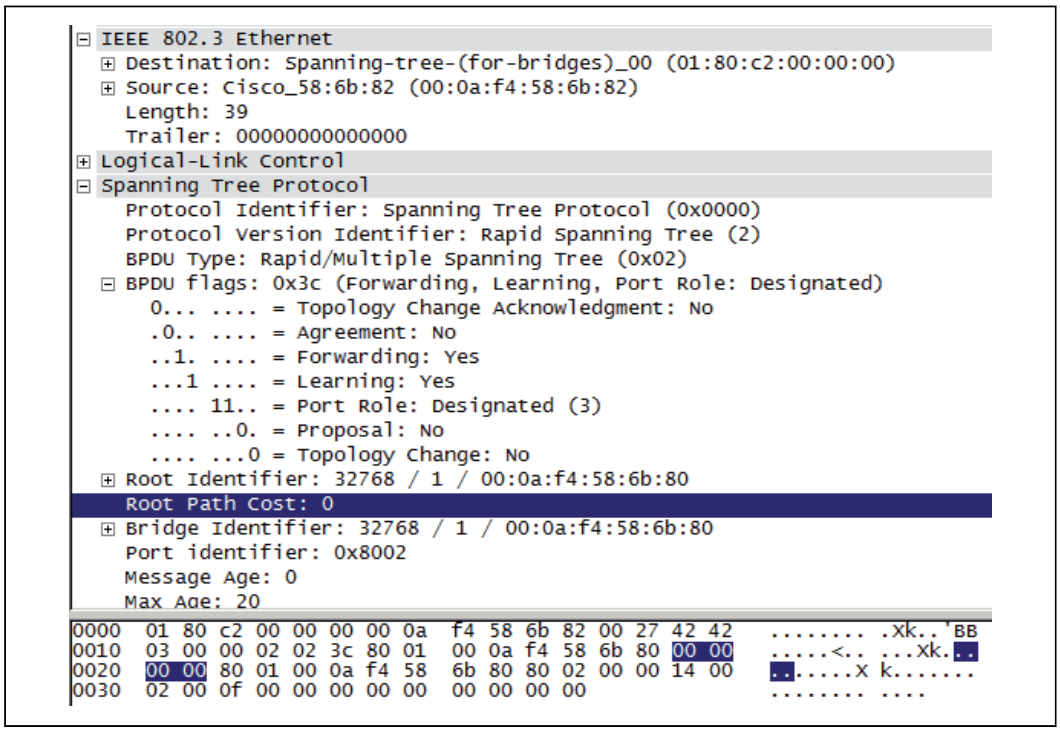
Statt einfach zum Forwarding zu wechseln und dann entweder zum Root- oder zu einem designierten Port zu werden, trennt RSTP diese beiden Ideen und reduziert die Anzahl der Zustände.

Tabelle 3-1: Vergleich der Port-Zustände

802.1D	802.1w
Disabled	Gestrichen
Blocked	Gestrichen
Listening	Gestrichen
Learning	Learning
Forwarding	Forwarding

Blocking wird zu einer Port-Rolle und unterteilt sich in Backup und alternativ blockierte Ports. Die designierten und Root-Rollen sind nun Variablen des Ports und werden zusammen mit der BPDU gesendet. Innerhalb des Flags-Feldes wird der Port-Status (Learning und Forwarding) und die Rollen mit der BPDU übertragen. Zusätzlich gibt es Änderungen in der Protokoll-Version und der Signalisierung. 802.1D-Nachrichten waren auf Konfiguration, TCN (Topology Change Notification) und TCN-ACKs beschränkt. RSTP fügt sogenannte Proposals (Vorschläge) und ACKs für diese Proposals hinzu. Ein Beispiel für diese Änderungen sehen Sie in Abbildung 3-32.

▼ Abbildung 3-32
Rapid Spanning Tree-Felder



Selbst mit diesen Änderungen ist ein Großteil des Protokolls gleich. Root-Ports sind immer noch diejenigen, die BPDUs empfangen und TCN-BPDUs senden (die zum Root-Switch zeigen), und designierte Ports senden immer noch BPDUs und empfangen TCN-BPDUs (die vom Root-Switch weg zeigen). Switch- und Port-Prioritäten werden ebenfalls auf die gleiche Weise eingesetzt.

Der Betrieb von RSTP

Im Gegensatz zum Spanning Tree nach 802.1D, bei dem Switches nur BPDUs senden, wenn sie Upstream welche empfange haben, senden RSTP-Switches BPDUs bei jedem Hello. Darüber hinaus wartet RSTP nicht 20 Sekunden, bis der Max-Age-Timer abläuft, sondern nur drei Hellos (6 Sekunden), bevor er Informationen von seinen Nachbarn als veraltet aussortiert. Das bedeutet, dass Fehler schneller erkannt werden. Die Hello- oder Konfigurations-BPDUs können also als eine Art »Keep Alive«-Nachricht betrachtet werden. RSTP erlaubt außerdem die sofortige Akzeptanz schlechterer BPDU-Informationen, falls die Verbindung zum Root-Switch verloren geht. Das ähnelt dem Verhalten von Backbonefast.

Ports werden nun ohne den Einsatz von Hersteller-Verbesserungen entsprechend ihrem Link-Typ identifiziert. Edge-Ports empfangen keine BPDUs und können daher direkt in den Forwarding-Status wechseln, ohne sich mit den anderen Port-Zuständen aufhalten zu müssen. Das ähnelt dem Verhalten von Portfast. Ein Edge-Port, der eine BPDU empfängt, wird zu einem Standard-Spanning-Tree-Port. Punkt-zu-Punkt-Links sind direkte Verbindungen zwischen Switches. Diese Ports wechseln auch schnell in den Forwarding-Zustand über, da hier eine Schleife weniger wahrscheinlich ist. Das basiert darauf, dass die Ports Vollduplex arbeiten. Diese Link-Typen können auch von Hand konfiguriert werden. Abbildung 3-33 zeigt die Ausgabe des `show spanning-tree`-Befehls nach der Aktivierung von RSTP. Beachten Sie die Änderung der Link-Typen.

Bei Punkt-zu-Punkt-Links vereinbaren Switches den Beginn des Forwardings über Proposals und entsprechende Zustimmung (Agreement). Bei Änderungen gehen Ports in den »Sync«-Zustand über, indem sie blockieren/ausrangieren oder zu Edge-Ports werden. Die Verhandlungen bestimmen die Root-Ports und schalten einige Ports direkt in den Forwarding-Zustand. Bei RSTP-Topologieänderungen starten Switches einen Topologieänderungs-Timer für alle designierten »Nicht-Edge«-Ports und den Root-Port. Spanning Tree-MAC-Adressen für diese Ports werden entfernt. Auf diese Weise werden die neuen Informationen schnell an die anderen Switches im

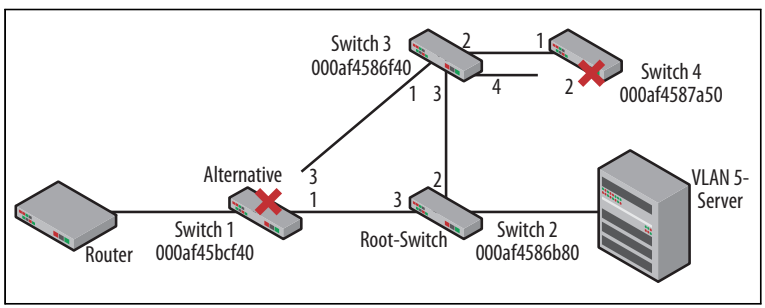
Netzwerk weitergeleitet. Die Konvergenzzeit wird drastisch reduziert. Das macht RSTP zu einem Teil von Redundanz- und Ausfallsicherungslösungen.

▼ **Abbildung 3-33**
Spanning Tree-Ausgabe für RSTP

```
Switch#sh spa
3w0d: %SYS-5-CONFIG_I: Configured from console by console

VLAN0001
  Spanning tree enabled protocol rstp
  Root ID     Priority    32769
              Address     000a.f458.6b80
              Cost        19
              Port        3 (FastEthernet0/3)
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    32769  (priority 32768 sys-id-ext 1)
              Address     000a.f458.6f40
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
              Aging Time 300

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Fa0/1            Desg FWD 19        128.1    P2p
Fa0/3            Root FWD 19        128.3    Shr
Fa0/5            Desg FWD 19        128.5    Edge P2p
```

Blockierte Alternativ- und Backup-Ports

Alternative Ports sind blockierte Ports, die immer noch BPDUs von anderen Bridges empfangen, obschon es bessere Pfade gibt. Auf der vorhin verwendeten Topologie aufbauend habe ich Switch 4 Downstream von Switch 3 eingefügt. Beim Verlust der »besseren« Verbindung zum Root-Switch über Switch 3 geht der *alternative* Port an Switch 1 schnell zum Forwarding über. Das ähnelt der Idee von Uplinkfast, geht aber wesentlich schneller.



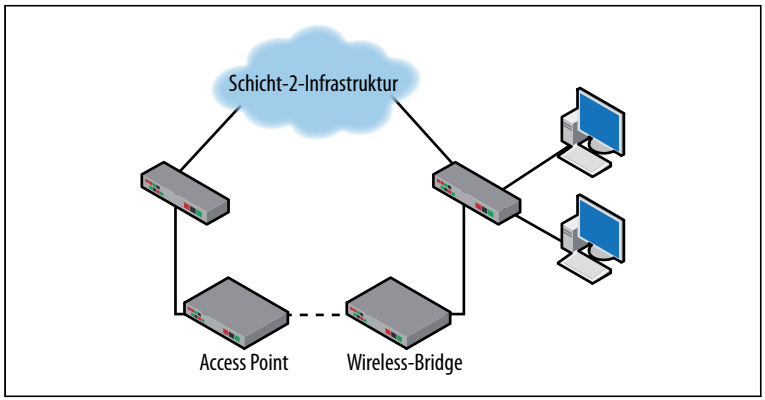
◀ **Abbildung 3-34**
Alternativ- und Backup-Ports

Backup-Ports empfangen BPDUs vom eigenen Switch, sind aber die »schlechteren« Ports. In diesem Fall werden sie blockiert und haben keine garantierte Route zurück zum Root-Switch. Das ist auch in Abbildung 3-34 bei Switch 4 zu sehen. Schauen wir uns Port 2 an Switch 4 genau an, erkennen wir, warum er blockiert wurde. Vergleicht man die BPDUs von Switch 3 nach Switch 4, dann ist die BPDU, die Port 4 verlässt, unterlegen. Sendet Port 2 an Switch 3 keine BPDUs mehr, sollte Port 4 an Switch 3 übernehmen. Doch Switch 4 wird der Zugriff auf den Root-Switch nicht garantiert, da sich der Gesamtpfad nicht verändert hat. Er läuft immer noch über Switch 3 und ist damit von der Konnektivität irgendwo im Netzwerk abhängig. Das ist einer der wenigen Fälle, in denen RSTP dem 802.1D-STP im Bezug auf die Konvergenz nicht überlegen ist.

Sicherheit

Sicherheitserwägungen bei Protokollen wie Spannung Tree zielen üblicherweise nicht in Richtung Datenverlust oder Einbruchsversuche. Vielmehr machen sich Administratoren Gedanken über die Unterbrechung des Netzwerkbetriebs und über DoS-Probleme (Denial of Service). Spanning Tree-Topologien lassen sich durch das Einspeisen von zusätzlichem Traffic relativ leicht lahmlegen. Die Verbindung mit einem beliebigen Port eines Switches bietet einen Vektor für die Einspeisung. Programme wie Yersinia ermöglichen es Angreifern, die notwendigen Pakete zu erzeugen. Solche Angriffe funktionieren, weil die Switches und Protokolle davon ausgehen, dass die eingehenden Pakete die richtigen Informationen enthalten. So könnte beispielsweise ein Angreifer in eine der in diesem Kapitel diskutierten Topologien eine »falsche« BPDU einschleusen. Die angreifende BPDU könnte eine Root-Priorität enthalten, die im Vergleich zum aktuellen Root-Switch sehr niedrig ist. Diese »falsche« BPDU erzwingt also eine neue Topologie, bei der alle Pfade in Richtung des neuen Root-Switches laufen. Sobald die Topologie konvergiert ist, kann der Angreifer die BDPUs entfernen und eine weitere Topologieänderung erzwingen. Solche Änderungen lassen den Netzwerk-Traffic in Richtung Angreifer fließen, d.h., die Benutzerdaten sind für den Angreifer potenziell sichtbar. Schutz vor solchen Angriffen bieten Befehle wie `root guard` und `bpdu guard`, die Angreifern das Einschleusen falscher BPDUs oder das Ersetzen des Root-Switches erschweren. Solche Angriffe können aber auch hausgemacht sein. Installiert ein Netzwerkadministrator einen Switch, ohne die Spanning Tree-Topologie zu berücksichtigen, kann das gleiche Szenario eintreten.

Eine weitere Erwägung für Netzwerkadministratoren ist das Drahtlos-Netzwerk. Viele Zugangspunkte (Access Points) können in Drahtlos-Bridges umgewandelt werden. Das macht man üblicherweise, um eine Netzwerkverbindung mit geografisch entfernten Sites oder Knoten herzustellen. Wie ein Zugangspunkt besitzt eine Drahtlos-Bridge ein kabelgebundenes und ein drahtloses Interface. Ein kleiner Ausrutscher kann zu einer Topologie wie in Abbildung 3-35 führen.



◀ Abbildung 3-35
Schleife beim Drahtlos-Bridging

Je nach Typ und Hersteller der Brigde nimmt diese am Spanning Tree teil – oder halt auch nicht. Stellen Sie sich vor, die Priorität der Drahtlos-Bridge ist kleiner als bei einem der verkabelten Switches. Die Topologie muss erneut konvergieren, und die Drahtlos-Bridge könnte zum Root-Switch werden. Drahtlos-Bridges haben nicht annähernd die Kapazität kabelgebundener Switches. Nimmt die Drahtlos-Bridge nicht am Spanning Tree teil, kann es zu einer nicht aufgelösten Schleife kommen oder dazu, dass ein Port in der Topologie blockiert wird. Das kann zu eingeschränkter Konnektivität für einen Teil des Netzwerks führen.

Lektüre

IEEE 802.1D Bridging-Standard: »IEEE Standard for Local and metropolitan area networks: Media Access Control (MAC) Bridges«, (berücksichtigt 802.1w), Jun. 2004.
IEEE 802.1Q VLAN-Standard: »IEEE Standards for Local and metropolitan area networks: Virtual Bridged Local Area Networks«, (berücksichtigt 802.1v und 802.1s), Mai 2006.
RFC 5556: »Transparent Interconnection of Lots of Links (TRILL): Problem and Applicability Statement«

Zusammenfassung

Spanning Tree ist ein Protokoll, das Teil nahezu jedes Schicht-2-Netzwerks ist. Da es standardmäßig läuft, wird es oft falsch verstanden oder ganz ignoriert. Dennoch übernehmen Spanning Tree und das schnellere Rapid Spanning Tree eine wichtige Aufgabe, da sie das Ethernet-Netzwerk vor Schicht-2-Schleifen schützen. Die in diesem Kapitel beschriebenen Szenarien und Datenpakete erläutern den Betrieb und das erwartete Verhalten dieses allgegenwärtigen Prozesses. Rapid-Spanning und 802.1D-Spanning teilen sich viele Eigenschaften, doch aufgrund der verbesserten Konvergenzgeschwindigkeit ist Rapid Spanning Tree für die Redundanz erste Wahl. Obwohl es die Protokolle schon eine Weile gibt und einige Funktionen von Routern übernommen wurden, geht die Arbeit an der Auflösung von Schleifen auf der 2. Schicht weiter. Projekte wie TRILL (Transparent Interconnection of Lots of Links) zeigen, dass die Leute, insbesondere Radia Perlman, mit dem Nachdenken über dieses Problem noch nicht fertig sind.

Fragen

1. In welchem Standard ist Spanning Tree definiert?
2. Die Hauptaufgabe von Spanning Tree besteht in der Eliminierung logischer Schleifen.
   a. WAHR
   b. FALSCH
3. Wie heißen die vier im Vergleichsalgorithmus verwendeten Felder?
4. Aus welchen Komponenten besteht das Root-ID-Feld?
5. Welche Ziel-MAC-Adresse wird in einer BPDU verwendet?
6. Beschreiben Sie den Unterschied zwischen einem Root-Port und einem designierten Port.
7. Welche Werte haben die Hello-, Max-Age- und Forward-Delay-Timer?
8. Nennen Sie drei von Cisco eingeführte Verbesserungen an Spanning Tree.
9. Rapid Spanning Tree ist für den Aufbau schneller, redundanter Netzwerke geeignet.
   a. WAHR
   b. FALSCH

10. Welche zwei Arten blockierter Ports werden von Rapid Spanning Tree definiert?

Antworten

1. 802.1D
2. WAHR
3. Root-ID, Pfadkosten, Bridge-ID, Port-ID
4. Bridge-Priorität und MAC-Adresse
5. 01:80:c2:00:00:00
6. Root-Ports weisen in Richtung Root-Switch, und designierte Ports weisen davon weg. Traffic fließt auf seinem Weg zum Root-Switch in die designierten Ports hinein und aus Root-Ports hinaus.
7. 15, 20 und 2 Sekunden
8. Portfast, Uplinkfast und Backbonefast
9. WAHR
10. Alternativ und Backup

Laborübungen

Übung 1: Capture einer BPDU

Material: Computer mit aktiver Netzwerkverbindung zu einem Switch und Packet-Capture-Software (Wireshark)

1. Falls es noch nicht geschehen ist, verbinden Sie den Switch mit der Netzwerkschnittstelle des Computers.
2. Sobald das Link-Licht grün wird, starten Sie die Packet-Capture-Software.
3. Untersuchen Sie die festgehaltenen Pakete, und finden Sie eine BPDU.
4. Öffnen Sie die BPDU, und untersuchen Sie die vom Vergleichsalgorithmus verwendeten Felder. Wie sehen die Werte aus?

Übung 2: BPDU-Adressanalyse

Material: Computer mit aktiver Netzwerkverbindung zu einem Switch und Paket-Capture-Software (Wireshark)

1. An der im ersten Experiment abgefangenen BPDU untersuchen wir die verwendeten Adressen. Suchen Sie nach Folgendem:
2. Quell- und Ziel-MAC-Adressen
3. Root-Bridge-MAC-Adresse
4. MAC-Adresse der sendenden Bridge
5. In welcher Beziehung stehen sie zueinander? Wie werden sie verwendet?

Übung 3: Switch an sich selbst zurückschleifen

Material: Managed Switch

1. Bevor Sie dieses Experiment starten, sollten Sie festlegen, was Sie tun, und versuchen herauszufinden, was mit dieser Schleife passiert.
2. Verbinden Sie mit einem Ethernet-Kabel einen Port des Switches mit einem anderen.
3. Was passiert mit den Link-Lichtern?
4. Welchen Status haben die Ports am verwalteten Interface? Sind welche blockiert? Seien Sie mit der Antwort sehr genau.

Übung 4: Switches in einer Schleife miteinander verbinden

Material: Zwei oder drei Managed Switches, ein Computer, der Pakete festhalten kann

1. Diese Übung verlangt Zugang zu einer Reihe von Switches. Zwei Switches reichen für diese Übungen aus.
2. Versuchen Sie, wie bei Übung 3, herauszufinden, was passiert, wenn die Schleife erzeugt wird. Seien Sie sehr genau in Bezug darauf, was passiert und warum.
3. Schalten Sie 2 oder 3 Switches in Reihe. Welcher Switch ist der Root-Switch? Wie sieht die BPDU an der am weitesten vom Root-Switch entfernten Stelle aus? Beachten Sie die vier Felder des Vergleichsalgorithmus.
4. Verbinden Sie die Switches zu einer Schleife. Welcher Port wird blockiert? Warum? Wie sehen die Änderungen der BPDUs für jedes Netzwerk-Segment aus?

Übung 5: Die Schleife entfernen

Material: Zwei oder drei Managed Switches, ein Computer, der Pakete festhalten kann

1. Entfernen Sie in der Topologie aus Experiment 4 die von ihnen aufgebaute physische Schleife.
2. Wie lange dauert es, bis der blockierte Port in den Forwarding-Status wechselt?
3. Was passiert, wenn Sie die Switch-Priorität an einem der Nicht-Root-Switches ändern? Wie lange dauert es, bis sich dieser Effekt in der Netzwerk-Topologie und den BPDUs widerspiegelt?

KAPITEL 4

VLANs und Trunking

In diesem Kapitel:

- Problem: Große Broadcast-Domains
- Was ist ein VLAN?
- Was ist ein Trunk?
- Erwägungen zum VLAN-Design
- Lektüre
- Zusammenfassung
- Fragen
- Antworten
- Laborübungen

Der Wechsel von einfachen Hubs hin zu vermittelten Netzwerken (Switched Networks) war eine große Verbesserung. Die Kontrolle über Kollisionen, erhöhter Durchsatz und von Switches zusätzlich angebotenen Features sind ein ausreichender Anreiz dafür, die Infrastruktur zu aktualisieren. Doch auf der zweiten Schicht vermittelnde Topologien haben auch so ihre Schwierigkeiten. Ausgedehnte flache Topologien können zu überfüllten Broadcast-Domains und zu Einschnitten in Bezug auf Sicherheit, Redundanz und Load-Balancing führen. Das kann durch den Einsatz virtueller lokaler Netzwerke (VLANs) gelindert werden. Dieses Kapitel behandelt die Struktur und den Betrieb von VLANs, wie sie in IEEE 802.1Q standardisiert sind. Die Diskussion umfasst auch Methoden der Bündelung (Trunking), die man zur Kopplung von Geräten in VLANs nutzt.

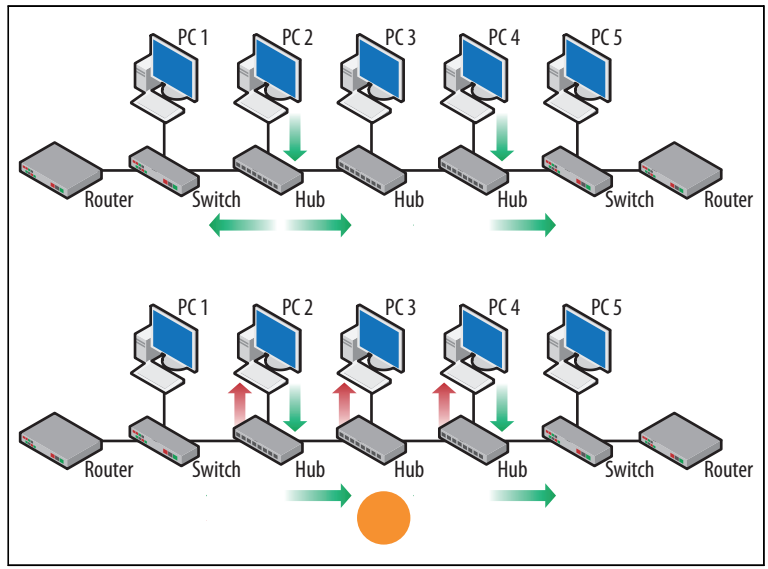
Problem: Große Broadcast-Domains

Innerhalb jedes einzelnen LAN-Segments werden Datenübertragungen durch das gesamte Segment propagiert. Erhöht sich der Datenverkehr, erhöht sich die Zahl der Kollisionen, und die sendenden Knoten müssen sich zurücknehmen und warten, bevor sie eine erneute Übertragung starten. Während die Kollision aufgelöst wird, müssen auch andere Knoten warten, was den Andrang im LAN-Segment weiter erhöht.

Die obere Seite von Abbildung 4-1 zeigt ein kleines Netzwerk, in dem PC 2 und PC 4 gleichzeitig versuchen, etwas zu senden. Die Frames werden von den Computern weg propagiert und kollidieren schließlich irgendwo zwischen den beiden Knoten auf der rechten Seite. Die erhöhte Spannung und Leistung propagiert dann vom Ort der Kollision weg. Beachten Sie, dass die Kollision nicht über die

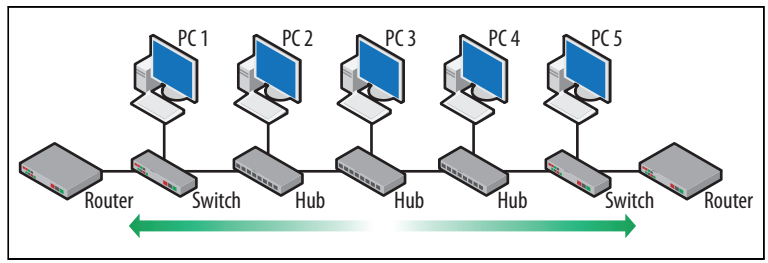
Switches an beiden Enden hinausgeht. Das sind die Grenzen der *Kollisionsdomäne* (Collision Domain). Das ist einer der Hauptgründe, warum Hubs durch Switches ersetzt werden. Hubs (und Access Points) skalieren einfach nicht besonders gut, wenn sich der Netzwerk-Traffic erhöht.

Abbildung 4-1 ▶
Vor und nach der Kollision

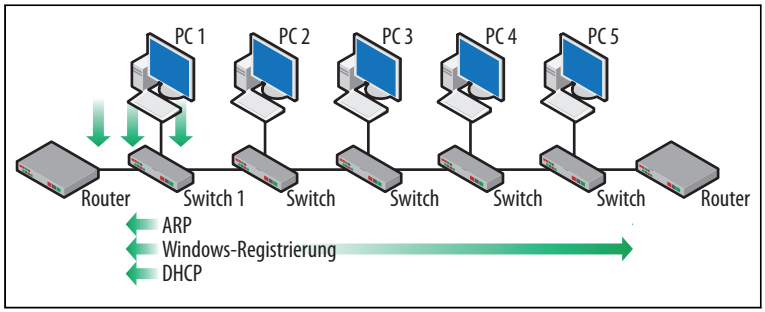


Die Verwendung von Switches in der zweiten Schicht eliminiert einen Großteil der Skalierungsprobleme, da solche Dinge wie Kollisionen herausgefiltert werden. Übertragungen werden stattdessen durch das Verhalten des Switches und der Broadcast-Domain bestimmt. Eine *Broadcast-Domain* definiert den Bereich, über den ein Broadcast-Frame propagiert wird. Zum Beispiel führt ein ARP-Request von PC 3 zu einem Broadcast-Frame, der durch alle Switches zu den Routern durchpropagiert wird (siehe Abbildung 4-2). Ein Broadcast-Frame verwendet die Broadcast-Adresse (`FF-FF-FF-FF-FF-FF`) als Ziel-MAC.

Abbildung 4-2 ▶
Broadcast-Domain



Durch die verbesserte Performance und die Filterung, die sich durch die Switches ergibt, ist die Versuchung groß, große Schicht-2-Topologien aufzubauen und viele Knoten einzufügen, doch das führt zu einer großen Broadcast-Domain. Das Problem besteht darin, dass alle Geräte im Netzwerk (Computer, Drucker, Switches etc.) Broadcast- und Multicast-Frames erzeugen, die die gesamte Broadcast-Domain durchlaufen und mit den Daten um Bandbreite kämpfen. Ein Großteil dieses Traffics dient zur Verwaltung des Netzwerks und umfasst Protokolle zur Adressauflösung (ARP), zur dynamischen Konfiguration von Hosts (DHCP), Spanning Tree (STP) und einer Mischung von Windows-Aufgaben. Abbildung 4-3 verdeutlicht das potenzielle Problem. Gehen Sie davon aus, dass PC1 die folgenden Requests generiert hat: ARP, Windows-Registrierung und DHCP.



◀ **Abbildung 4-3**
Anstieg der Broadcast-Frames

Da alle Requests einen Broadcast-Frame nutzen, werden die Frames in alle Richtungen weitergeleitet, sobald sie Switch 1 erreichen. Da die anderen Switches in der Topologie dem prompt folgen, laufen die Frames durch das gesamte Netzwerk und werden von allen Knoten und Routern empfangen.

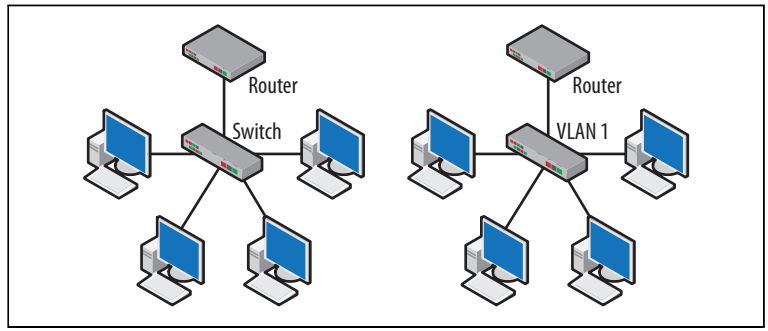
Erhöht sich die Zahl der Netzwerk-Knoten, steigt auch der Overhead. Jeder Switch kann mit Dutzenden Knoten verbunden sein, von denen jeder mehrere Broadcast-Frames erzeugt. Wird genug Traffic erzeugt, kann auch ein vermitteltes Netzwerk eine schlechte Performance aufweisen. Der Einsatz von VLANs kann dieses Problem lösen, indem er die Broadcast-Domain und damit den Traffic aufteilt.

Was ist ein VLAN?

Ein virtuelles lokales Netzwerk (Virtual Local Area Network, VLAN) ist eine logische Gruppierung von Ports unabhängig von deren Lage. Ein einzelnes VLAN (und die darin angeschlossenen Knoten) verhal-

ten sich wie ein separates Schicht-3-Netzwerk. Die Zugehörigkeit zu einem VLAN muss nicht auf eine Reihe aufeinanderfolgender Ports oder gar auf Ports an einem Switch begrenzt sein. Abbildung 4-4 zeigt einen sehr weit verbreiteten Anwendungsfall, bei dem die Knoten mit einem Switch und der Switch mit einem Router verbunden ist. Sieht man sich die linke Seite an, geht man automatisch davon aus, dass alle Knoten im gleichen IP-Netzwerk liegen, da sie mit dem gleichen Router-Interface verbunden sind.

Abbildung 4-4 ▶
Einfache Switch- und VLAN-Topologie

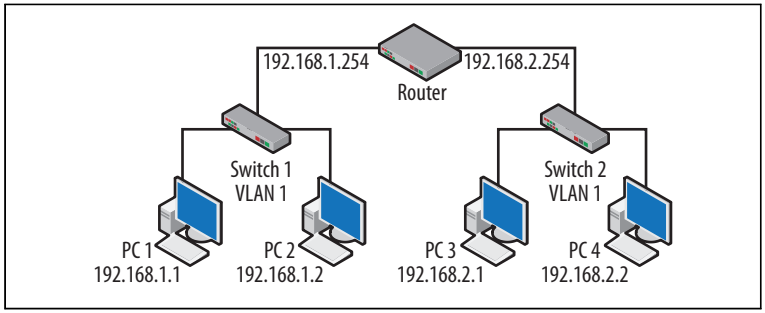


Was bei der Topologie auf der linken Seite nicht offensichtlich wird, ist die Tatsache, dass alle Knoten Teil des gleichen VLANs sind. Ein anderer Blick auf diese Topologie basiert daher auf dem VLAN auf der rechten Seite. Bei Cisco-Geräten ist beispielsweise VLAN 1 das Standard-VLAN. Diese Ausgangskonfiguration schließt alle Ports als Mitglieder ein, was sich in der Quelladresstabelle (Source Address Table, SAT) widerspiegelt. Sie wird häufig als die Tabelle beschrieben, die Frames basierend auf der Ziel-MAC-Adresse an den richtigen Schicht-2-Port weiterleitet. Mit der Einführung von VLANs spiegelt die SAT die Port-zu-MAC-Abbildung auf VLAN-Basis wider, was zu fortgeschrittenen Forwarding-Entscheidungen führt. Abbildung 4-5 zeigt die Ausgabe der `show mac-address-table`- und `show vlan`-Befehle. Alle Ports (Fa0/1 bis Fa0/24) liegen in VLAN 1.

```
Switch#sh mac-address-table
          Mac Address Table
-------------------------------------------

Vlan    Mac Address       Type        Ports
----    -----------       --------    -----
 All    000c.85aa.ea40    STATIC      CPU
 All    0100.0ccc.cccc    STATIC      CPU
 All    0100.0ccc.cccd    STATIC      CPU
 All    0100.0cdd.dddd    STATIC      CPU
   1    000a.f458.6c58    DYNAMIC     Fa0/24
   1    0013.f7d1.de9b    DYNAMIC     Fa0/24
   1    0013.f7d1.e016    DYNAMIC     Fa0/1
   1    0013.f7d2.06d5    DYNAMIC     Fa0/24
   1    0013.f7d2.06e1    DYNAMIC     Fa0/2
Total Mac Addresses for this criterion: 9
Switch#
Switch#show vlan

VLAN Name                             Status    Ports
---- -------------------------------- --------- -------------------------------
1    default                          active    Fa0/1, Fa0/2, Fa0/3, Fa0/4
                                                Fa0/5, Fa0/6, Fa0/7, Fa0/8
                                                Fa0/9, Fa0/10, Fa0/11, Fa0/12
                                                Fa0/13, Fa0/14, Fa0/15, Fa0/16
                                                Fa0/17, Fa0/18, Fa0/19, Fa0/20
                                                Fa0/21, Fa0/22, Fa0/23
```

▲ Abbildung 4-5
Switch-SAT- und VLAN-Ausgabe

Eine weitere typische Topologie sehen Sie in Abbildung 4-6, bei der zwei Switches durch einen Router getrennt sind. In diesem Fall ist eine Gruppe von Knoten mit jedem Switch verbunden. Die Knoten eines Switches verwenden das gleiche IP-Adressierungsschema. Es gibt zwei Netzwerke: 192.168.1.0 und 192.168.2.0.

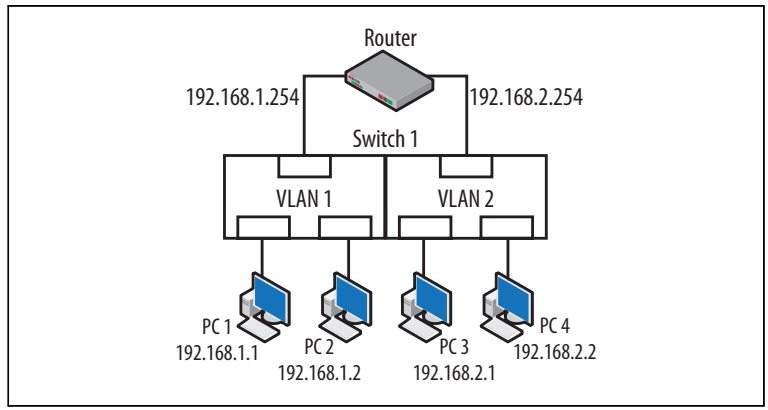


◀ Abbildung 4-6
Router, Switch und VLANs

Beachten Sie, dass beide Switches im gleichen VLAN liegen, da die Switches eines Herstellers das gleiche Nummerierungsschema verwenden, solange keine Konfigurationsänderungen vorgenommen wurden. Ausgehender Netzwerk-Traffic muss zum Forwarding an den Router gesendet werden. Router geben Schicht-2-Unicast-, -Multicast- und -Broadcast-Frames nicht weiter. VLANs bieten eine ver-

gleichbare logische Topologie, bei der die Knoten innerhalb des VLANs ein gemeinsames Adressierungsschema nutzen und ausgehenden Traffic zur Weiterleitung an den Router schicken. Indem man ein zusätzliches VLAN auf einem der Switches einrichtet und das andere entfernt, kann man die Topologie aus Abbildung 4-6 neu anordnen, wie es in Abbildung 4-7 zu sehen ist.

Abbildung 4-7 ▶
Ein Switch, mehrere VLANs



Ein VLAN arbeitet wie ein IP-basiertes Schicht-3-Netzwerk. Knoten im 192.168.1.0-Netzwerk müssen also den Router nutzen, wenn sie mit Knoten im 192.168.2.0-Netzwerk kommunizieren wollen, *obwohl alle Computer an den gleichen Switch angeschlossen sind*. Um zwischen VLANs kommunizieren zu können, muss die Routingfunktionalität Teil der Topologie sein. Schicht-2-Unicast-, -Multicast- und Broadcast-Traffic überschreitet die Grenzen des VLANs nicht, weshalb in VLAN 1 generierter Traffic für die Knoten in VLAN 2 nicht sichtbar ist. Nur der Switch ist sich der VLANs bewusst. Die Knoten und die Routen »wissen« nicht, dass VLANs verwendet werden. Durch das Hinzufügen der Routingentscheidung kann die Schicht-3-Funktionalität nur für zusätzliche Sicherheitseinstellungen, für die Eindämmung von Problemen und Traffic sowie für das Load-Balancing genutzt werden.

Auswirkungen von VLANs

Durch die Konfiguration mehrerer VLANs auf einem Switch wird die Größe der jeweiligen Broadcast-Domain reduziert. Der Overhead-Traffic ist geringer und reduziert den Bandbreiten-Kampf mit dem eigentlichen Datenverkehr. Anders ausgedrückt: Ein Knoten in einem bestimmten VLAN muss gegen weniger Broadcast-Traffic ankämpfen. Da das Forwarding-Verhalten des Switches auf der MAC-Adresse der SAT basiert, gelten die folgenden Regeln:

- Bei bekannten Unicast-Zielen leitet der Switch den Frame nur an den Zielport weiter.
- Bei unbekannten Unicast-Zielen leitet der Switch den Frame an alle aktiven Ports mit Ausnahme des Ursprungsports weiter. Das wird als Flooding (Fluten) bezeichnet.
- Bei Multicast- und Broadcast-Zielen leitet der Switch den Frame an alle aktiven Ports mit Ausnahme des Ursprungsports weiter.

Allerdings muss der Switch nun zusätzlich noch das VLAN des Zielknotens berücksichtigen. Wenn PC1 in Abbildung 4-7 einen ARP-Request sendet, leitet der Switch den Frame nicht einfach an alle Ports weiter, sondern erkennt, dass dieser Frame aus VLAN 1 stammt. Daher sieht nur PC2 auf der ganz linken Router-Schnittstelle (192.168.1.254) den Frame.

Ziele und Vorteile aus dem 802.1Q-Standard :

- VLANs werden von allen IEEE 802-LAN-MAC-Protokollen unterstützt, über Shared-Media-LANs ebenso wie über Punkt-zu-Punkt-LANs.
- VLANs ermöglichen die einfache Administration logischer Gruppenarbeitsplätze. Diese können miteinander so kommunizieren, als würden sie im gleichen LAN liegen. Sie ermöglichen auch eine einfachere Verwaltung von Verschiebungen, Ergänzungen und Änderungen der Mitglieder dieser Gruppen.
- Der Datenverkehr zwischen VLANs ist beschränkt. Switches leiten Unicast-, Multicast- und Broadcast-Traffic nur in LAN-Segmente des VLANs, zu dem dieser Traffic gehört.
- Soweit es möglich ist, erhalten VLANs die Kompatibilität mit existierenden Switches und Endgeräten aufrecht.
- Sind alle Switch-Ports so konfiguriert, dass sie »ungetaggte« Frames senden und empfangen (Frames von/an nicht-VLAN-fähige Geräte), dann arbeiten die Switches im ISO/IEC 15802-3 Plug-and-Play-Modus. Endgeräte können über das Bridged LAN kommunizieren.

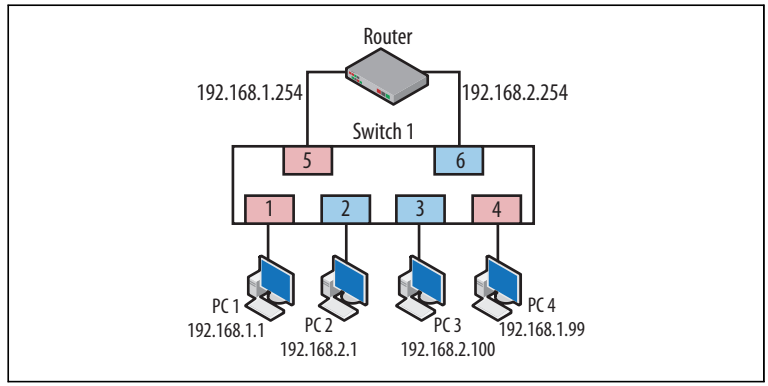
VLAN-Ports müssen nicht zusammenhängen

Da VLANs logische Gruppen von Knoten darstellen, die vom Ort unabhängig sind, spielt es keine Rolle, wo man die Knoten verbindet. Abbildung 4-8 demonstriert das Konzept. Die Topologie in Abbildung 4-7 wird mit den IP-Adressen der geänderten Netzwerkknoten abgebildet. Um die Dinge besser unterscheiden zu können, wird hier VLAN 1 rot und VLAN 2 blau dargestellt. Die Ports 1, 4

und 5 sind Teil des roten VLAN 1, während die Ports 2, 3 und 6 Teil des blauen VLAN 2 sind.

Meistens ist es so, dass die Netzwerk-Techniker die Topologie nicht jedes Mal neu verdrahten wollen, wenn ein neuer Knoten angeschlossen wird. Also wird der Host einfach mit einem freien Port verbunden, und dieser Port wird dem entsprechenden VLAN zugeordnet. Die Idee ist einfach, dass das Verhalten das gleiche ist, unabhängig davon, ob die Ports direkt nebeneinander liegen oder nicht. Daher können PC1 und PC4 direkt miteinander kommunizieren, müssen aber den Router nutzen, um an PC2 und PC3 zu gelangen. Die im roten VLAN 1 gesendeten Frames sind im blauen VLAN 2 nicht zu sehen.

Abbildung 4-8 ▶
Nicht zusammenhängende VLANs

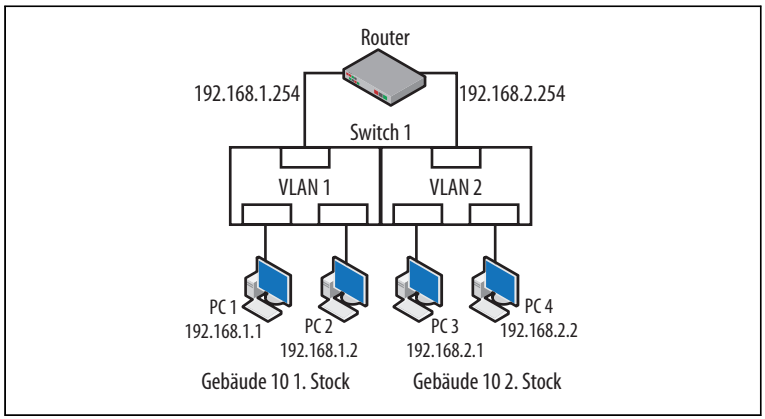


Arten von VLANs

Es gibt zwei Arten von VLANs: statische und dynamische. Mit beiden Arten lassen sich kleine und große geografische Gebiete abdecken. Die bislang diskutierte Art von VLAN (ein in mehrere VLANs unterteilter Switch) wird als statisches VLAN bezeichnet. Die Mitgliedschaft wird größtenteils durch die geografische Lage bestimmt und dadurch, mit welchem Port ein bestimmter Knoten verbunden ist. Die meisten Knoten eines bestimmten VLANs befinden sich üblicherweise im gleichen Gebäude, Gang oder in einer Reihe von Büros. Bei solchen VLANs kann man auch von einer lokalen Mitgliedschaft sprechen.

Abbildung 4-9 zeigt ein Beispiel dafür, wie Knoten und VLANs angeordnet sein können. PC1 und PC2 liegen physisch im gleichen Gebäudeteil und werden daher dem gleichen VLAN zugeordnet. Das Gleiche gilt für PC3 und PC4. Es ist anzunehmen, dass sie von Benutzern der gleichen Abteilung verwendet werden. Diese Art der

Topologie wird von einem Netzwerkadministrator von Hand konfiguriert. Er weist die Ports eines Switches einem bestimmten VLAN zu. Abermals wissen Knoten und Router nichts von den VLANs.

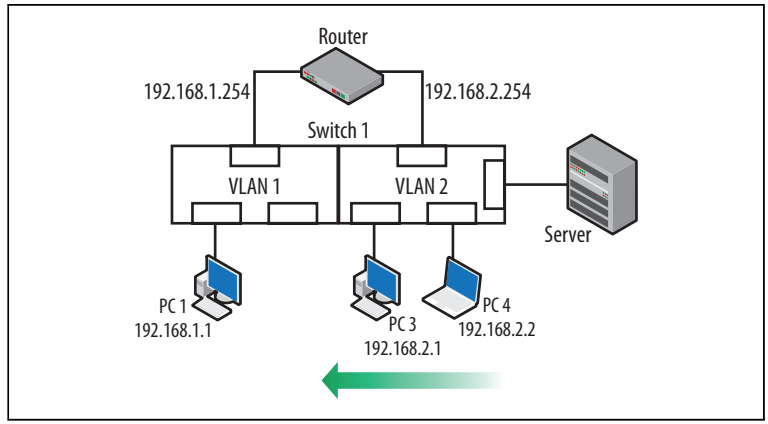


◀ **Abbildung 4-9**
Statisches VLAN, lokale Mitgliedschaft

Die meisten VLANs sind mit statischer Mitgliedschaft konfiguriert. Bei Topologien, wie sie oben beschrieben wurden, bleiben die Knoten mit dem gleichen Port verbunden, und es gibt keine Notwendigkeit, die VLAN-Mitgliedschaft zu ändern. Der Computer ist üblicherweise an einen Schreibtisch oder Arbeitsplatz gebunden, der einem Mitarbeiter zugeordnet ist, und darum muss man sich keine Gedanken darüber machen, dass die Maschine ihre Position ändert.

Es gibt aber Fälle, in denen die Knoten sich bewegen. Es könnte notwendig sein, auf unterschiedliche Ressourcen zuzugreifen. Ports könnten zu unterschiedlichen Zeiten von verschiedenen Abteilungen benutzt werden, oder es könnten unterschiedliche Sicherheitsgrade benötigt werden. Dynamische VLANs sind für diese Fälle besser geeignet. Bei dynamischen VLANs können sich die Knoten bewegen, ohne dass sich die VLAN-Mitgliedschaft ändert. Man schließt sich also einfach an, einen Port an und der Switch konfiguriert den Port automatisch für die Mitgliedschaft im richtigen VLAN. Ein Port, der bei Knoten A für den Zugriff auf VLAN 1 konfiguriert war, würde dann für Knoten B in das VLAN 2 wechseln. Betrachten wir den Fall in Abbildung 4-10. PC4 ist nun ein Laptop und wechselt von einem Port in VLAN 2 zu einem Port in VLAN 1.

Abbildung 4-10 ▶
Von einem VLAN zum anderen wechseln



1. Fall: DHCP

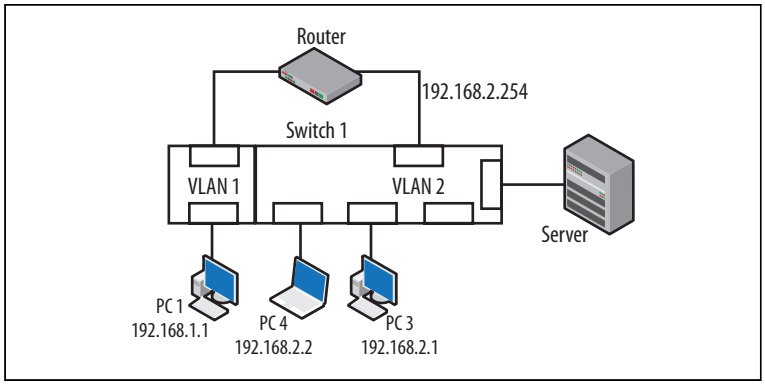
Wird DHCP eingesetzt, erhält PC4 beim Wechsel einfach eine neue IP-Adresse (auch wenn dies nicht garantiert wird). Das ist das übliche Verhalten von Knoten, die die Verbindung zu einem Netzwerk in einem bestimmten VLAN herstellen. Sind allerdings bestimmte Dienste oder Sicherheitsmaßnahmen aktiv und verlangt das Unternehmen eine Trennung der VLANs, dann kann diese Konfiguration ein Problem mit sich bringen: den Zugriff auf den Server. Im neuen Netzwerk ist PC4 möglicherweise nicht mehr in der Lage, den richtigen Server zu erreichen, oder es ist eine zusätzliche Konfiguration notwendig, um den Wechsel zu unterstützen.

2. Fall: Kein DHCP

Ist die IP-Adresse von PC4 statisch konfiguriert, passt sie beim Wechsel nicht mit zum neuen Netzwerk. Er kann die IP-Adresse des Gateways oder Servers nicht mehr erreichen. In diesem Fall hat der Knoten keinerlei Verbindung mehr.

Lösung: Dynamische VLANs

Ist der Switch nun clever genug, zu erkennen, dass PC4 jetzt an einem neuen Port hängt, kann er die Verbindung möglicherweise automatisch reparieren. Sobald PC4 mit dem neuen Port verbunden ist, generiert er Traffic. Empfängt der Switch ein Frame von PC4, führt er einen Datenbank-Lookup durch, um die VLAN-Mitgliedschaft zu ermitteln und dem Port das passende VLAN zuzuweisen. Sobald das geschehen ist, kann PC4 wie vor dem Umzug kommunizieren. Die neue Topologie ist in Abbildung 4-11 zu sehen. Der Knoten muss nicht einmal seine IP-Adresse ändern.

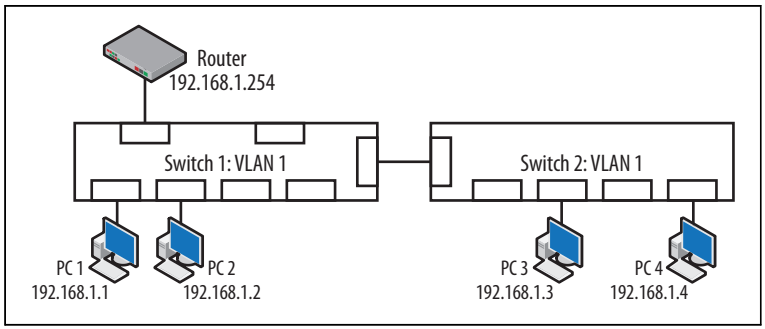


◀ **Abbildung 4-11**
Neue dynamische VLAN-Topologie

Doch woher weiß der Switch das? Die am weitesten verbreitete Methode der Zuweisung dynamischer VLAN-Mitgliedschaften nutzt die MAC-Adresse. Sobald der Knoten seinen ersten Frame erzeugt, schließt der Switch die Abfrage der MAC-Adresse ab und weist den Port zu. Die Knoten wissen immer noch nicht, dass mit VLANs gearbeitet wird. Die VLAN-Mitgliedschaft kann auch anhand anderer Kriterien vergeben werden oder an Authentifizierungsschemata wie 802.1X gekoppelt sein.

VLANs zwischen Switches

Bisher haben wir den Einsatz von VLANs auf einem einzelnen Switch betrachtet. Es stellt sich natürlich die Frage, was passiert, wenn das Gesamtnetzwerk aus mehreren Switches besteht. Wie funktioniert das? Die Antworten hängen von der Konfiguration der Switches ab. Eine Standardkonfiguration ist in Abbildung 4-12 zu sehen. Man erkennt zwei miteinander verbundene Switches, an die einfach einige Knoten angeschlossen sind. Das Standard-VLAN ist (wenn wir von Cisco-Geräten ausgehen) für beide Switches VLAN 1. Das bedeutet auch, dass die Verbindungen zwischen den Switches ebenfalls in VLAN 1 liegen. Der Router stellt für alle Knoten den Ausgang dar.



◀ **Abbildung 4-12**
Mehrere Switches, ein VLAN

Bei dieser Standard-Topologie haben die Knoten keine Schwierigkeiten, die Verbindung untereinander herzustellen, weil die SATs auf den Switches zeigen, dass alle im gleichen VLAN liegen. Der Unicast-, Multicast- und Broadcast-Traffic kann sich also frei bewegen. Beachten Sie auch, dass die Knoten im gleichen IP-Netzwerk liegen. Die Verbindung zwischen den Switches übernimmt entweder ein gekreuztes Kabel oder ein Uplink-Port.

Probleme treten auf, wenn neue VLANs angelegt werden, wie in Abbildung 4-13 zu sehen ist. Da die VLANs Schicht-3-Grenzen um die mit den Hosts verbundenen Ports ziehen, können diese nicht mehr kommunizieren.

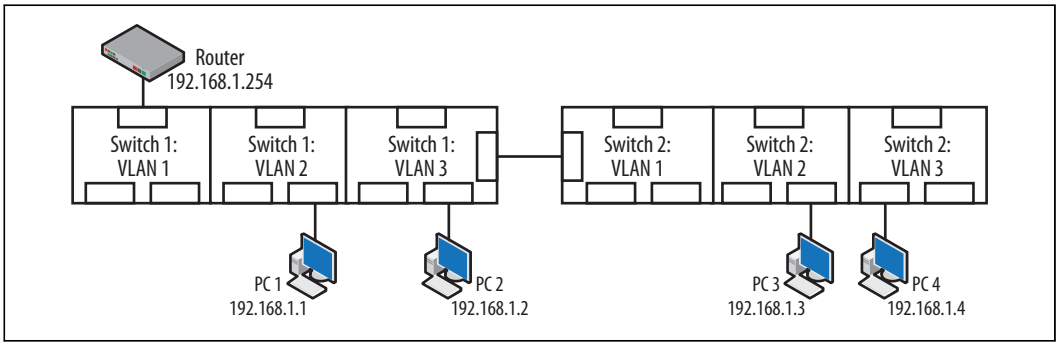
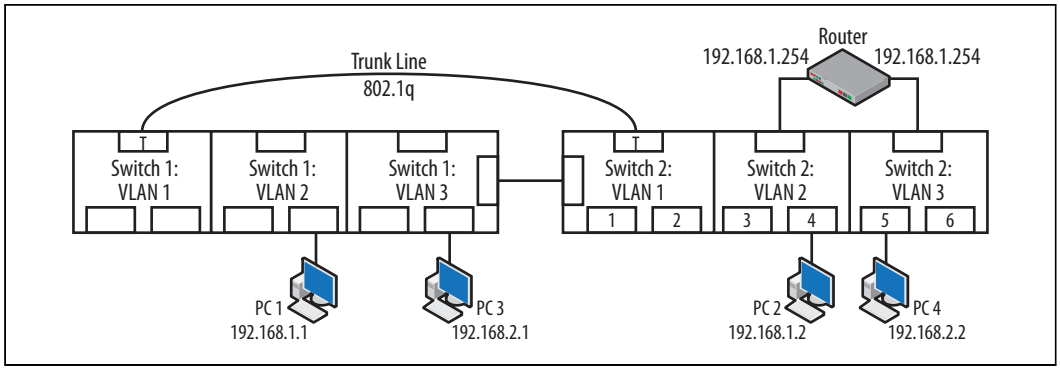


Abbildung 4-13 ▲ Probleme mit zusätzlichen VLANs

Sieht man sich Abbildung 4-13 an, erkennt man eine Reihe von Problemen. Erstens liegen alle Computer im gleichen IP-Netzwerk, obwohl sie an unterschiedliche VLANs angeschlossen sind. Zweitens ist der Router von allen Knoten abgeschnitten, da er in VLAN 1 liegt. Und schließlich sind die Switches über verschiedene VLANs miteinander verbunden. Jedes Problem würde für sich genommen die Kommunikation erschweren, aber zusammengenommen gibt es nur wenig oder gar keine Kommunikation im Netzwerk.

Es kommt häufig vor, dass ein Switch voll ist oder dass Knoten innerhalb der gleichen administrativen Einheit geografisch voneinander getrennt sind. In diesen Fällen kann ein VLAN über eine sogenannte Trunk Line auf benachbarte Switches ausgedehnt werden. Wir gehen im folgenden Abschnitt noch genauer auf Trunks ein. Im Moment reicht es, wenn Sie wissen, dass separate Switches verbindende Trunks (unter anderem) VLAN-Informationen zwischen Netzwerkgeräten übertragen können. Abbildung 4-14 schlägt verschiedene Änderungen vor, um die in Abbildung 4-13 gezeigten Punkte zu korrigieren.



▲ **Abbildung 4-14**
Mittels Trunking reparierte Topologie

Die Korrekturen an der Topologie sind wie folgt:

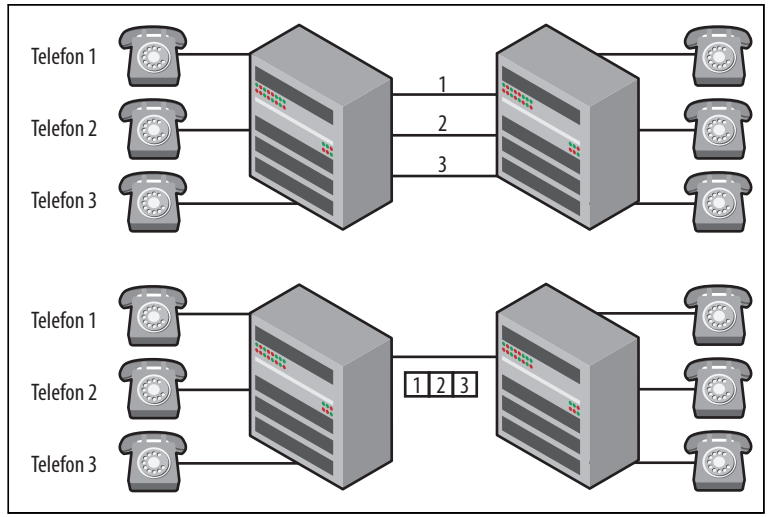
- PC1 und PC2 wurden dem Netzwerk 192.168.1.0 und VLAN 2 zugeordnet.
- PC3 und PC4 wurden dem Netzwerk 192.168.2.0 und VLAN 3 zugeordnet.
- Die Router-Interfaces sind mit den VLANs 2 und 3 verbunden.
- Die Switches sind über Trunk Lines miteinander verbunden.

Beachten Sie, dass die Trunk-Ports scheinbar in VLAN 1 liegen. Der Buchstabe T deutet aber an, dass das nicht der Fall ist. Trunk-Ports liegen in keinem bestimmten VLAN. Da sich die VLANs nun über mehrere Switches erstrecken, können die Knoten physisch an einem beliebigen Ort liegen und dennoch Mitglied des gleichen VLANs sein. Wenn mehrere Switches mit VLANs konfiguriert sind und die Ports sich um die VLAN-Mitgliedschaft kümmen, bezeichnet man die Architektur als »Ende-zu-Ende« und »statisch«. Es ist nicht ungewöhnlich, dass diese Switches in verschiedenen Schränken oder auch in unterschiedlichen Gebäuden liegen. Im gleichen Schrank liegende Switches können ebenfalls über Trunk Lines verbunden werden.

Was ist ein Trunk?

Ganz allgemein kann man eine Trunk Line auf zwei Arten betrachten. In der Telefonie steht der Begriff *Trunk Line* für die Amtsleitung, die die Büros oder die Verteiler verbindet. Diese Verbindungen repräsentieren eine erhöhte Anzahl von Leitungen oder Zeitmultiplex-Leitungen, wie in Abbildung 4-15 zu sehen ist.

Abbildung 4-15 ▶
Telefonleitungen und Trunks

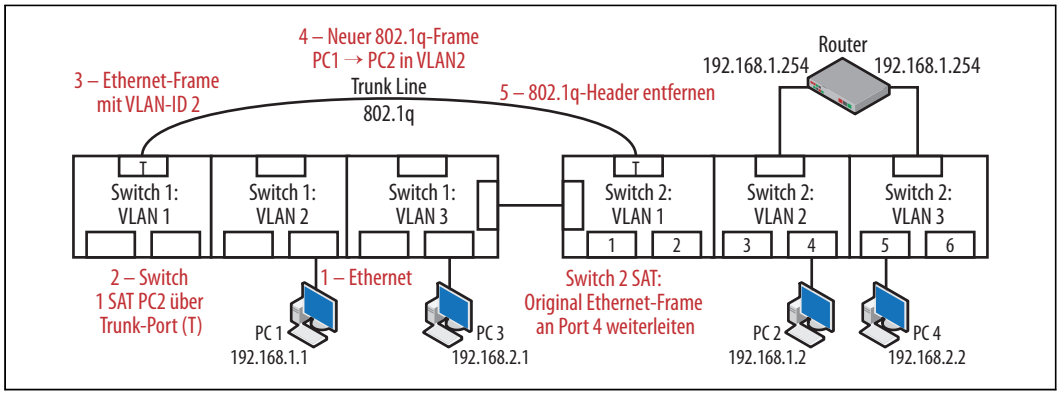


Bei Datennetzen haben Trunks nur wenig mit der Erhöhung der Verbindungsanzahl zwischen den Switches zu tun. Die primäre Aufgabe einer Trunk Line in einem Datennetz besteht darin, VLAN-Informationen zu befördern. Die Trunk Line in Abbildung 4-14 transportiert VLAN- und Quality-of-Service-Informationen für den teilnehmenden Switch.

Ist eine Trunk Line installiert, wird ein Trunking-Protokoll verwendet, um die Ethernet-Frames zu modifizieren, während sie durch die Trunk Line laufen. Bei den Ports in Abbildung 4-14, die die Switches miteinander verbinden, handelt es sich um Trunk Ports. Das bedeutet auch, dass es für Switches mehr als einen Betriebsmodus gibt. Standardmäßig werden alle Ports als *Zugriffs-Ports* (Access Ports) bezeichnet. Das beschreibt einen Port, der von einem Computer oder einem anderen Endgerät genutzt wird, um auf das Netzwerk »zuzugreifen«. Wird ein Port genutzt, um Switches zu verbinden und VLAN-Informationen zu transportieren, wird er im Trunk-Modus betrieben. Bei einem Cisco-Switch würden Sie beispielsweise den Befehl *mode* verwenden, um die Änderung vorzunehmen. Geräte anderer Hersteller zeigen an, dass der Port nun »getaggt« ist, d.h., dass nun eine VLAN-ID in die Frames eingefügt wird. Der 802.1Q-Standard umfasst auch Vorkehrungen für *Hybrid-Ports*, die sowohl getaggte als auch ungetaggte Frames verstehen. Um es deutlich zu machen: Knoten und Router sind sich der VLANs häufig nicht bewusst und verwenden »untaggte« Standard-Ethernet-Frames. Trunk Lines, die VLAN- oder Prioritätswerte liefern, verwenden »getaggte« Frames. Ein Beispiel für einen getaggten Frame sehen Sie in Abbildung 4-17.

Bei Trunk-Ports läuft also ein Trunking-Protokoll, das es erlaubt, VLAN-Informationen in jeden Frame einzufügen, der durch die Trunk Line läuft. Für die Konfiguration sind generell zwei Schritte notwendig: Umschaltung des Ports in den Trunk-Modus und Ermittlung der zu verwendenden Kapselung (des Trunking-Protokolls).

Mit Abbildung 4-16 wollen wir ein Beispiel durchgehen, wie zwei Knoten über eine Trunk Line miteinander kommunizieren. Dieser Prozess besteht aus mehreren Schritten (zusätzlich zum Host-Routing), weshalb Abbildung 4-16 basierend auf den Schritten gekennzeichnet ist.



▲ **Abbildung 4-16**
Trunking-Traffic zwischen Switches

PC1 sendet Daten an PC2, nachdem dessen Host-Routingtabelle verarbeitet wurde. Diese Knoten liegen im gleichen VLAN, sind aber an unterschiedliche Switches angeschlossen. Hier der grundlegende Prozess:

1. Der Ethernet-Frame verlässt PC1 und wird von Switch 1 empfangen.
2. Die SAT von Switch 1 besagt, dass das Ziel am anderen Ende der Trunk Line liegt.
3. Switch 1 verwendet das Trunking-Protokoll, um den Ethernet-Frame um die VLAN-ID zu ergänzen.
4. Der neue Frame verlässt Switch 1 über den Trunk-Port und wird von Switch 2 empfangen.
5. Switch2 liest die VLAN-ID ein und entfernt das Trunking-Protokoll.
6. Der Original-Frame wird basierend auf der SAT von Switch 2 an sein Ziel (Port 4) weitergeleitet.

Das Paket in Abbildung 4-17 zeigt Details dieser Modifikation. In diesem Fall wurde IEEE 802.1Q als Trunking-Protokoll verwendet. Der Frame ist ein ICMP Echo-Request von PC1→PC2. Da er durch die Trunk Line läuft, muss der VLAN-Tag eingefügt werden, damit Switch 2 weiß, wie er das Paket weiterleiten soll.

```
Ethernet II, Src: SmcNetwo_d1:dd:3d (00:13:f7:d1:dd:3d), Dst: SmcNetwo_d2:07:98 (00:13:f7:d2:07:98)
802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 2
  000. .... .... .... = Priority: Best Effort (default) (0)
  ...0 .... .... .... = CFI: Canonical (0)
  .... 0000 0000 0010 = ID: 2
  Type: IP (0x0800)
Internet Protocol, Src: 192.168.1.1 (192.168.1.1), Dst: 192.168.1.2 (192.168.1.2)
Internet Control Message Protocol
```

Abbildung 4-17 ▲ Ethernet-Frame mit 802.1Q-Trunking

Der Ethernet-Frame ist intakt, doch es gibt verschiedene zusätzliche Felder wie die VLAN-ID. In diesem Fall liegen die beiden kommunizierenden Computer in VLAN 2. Der Binärwert 0000 0000 0010 ist zu sehen. Beachten Sie, dass die IP- und ICMP-Header nicht verändert wurden. Da dies aber eine Änderung an einem realen Frame ist, muss der Cyclical Redundancy Check (CRC) am Ende des Ethernet-Frames neu berechnet werden. Dem Trunking wird üblicherweise nicht so viel Aufmerksamkeit gewidmet, wie es verdient, doch sobald auf den Switches VLANs konfiguriert sind, muss ein Trunking-Protokoll verwendet werden, wenn die VLANs von einem Switch zum nächsten erhalten bleiben sollen. Ohne Trunk liegen die Knoten wahrscheinlich alle im gleichen VLAN, was zu den vorhin beschriebenen Problemen führen kann. Trunks und VLANs sind unverzichtbarer Bestandteil von Standard-Topologien.

Trunking-Protokoll-Standards

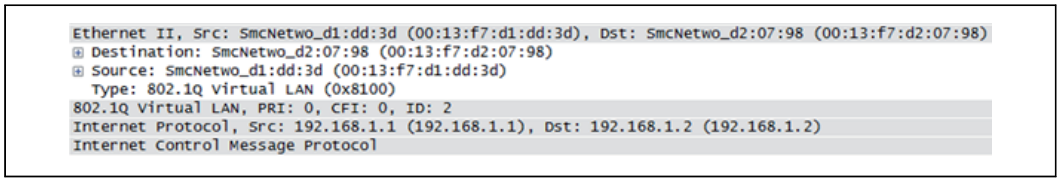
Zwei Trunking-Protokolle werden in modernen Kommunikationsnetzwerken eingesetzt: Inter-Switch Link (ISL) von Cisco und das bereits erwähnte, nicht-proprietäre IEEE 802.1Q. Von den beiden ist IEEE 802.1Q der Industrie-Standard. Selbst Cisco-Switches verwenden nun standardmäßig IEEE 802.1Q (dot1q).

IEEE 802.1Q

Der IEEE 802.1Q-Standard heißt vollständig »IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks« und beschäftigt sich hauptsächlich mit VLANs selbst. Das Trunking-Protokoll, also das »Tagging« der Frames, wird in späteren Abschnitten von 802.1Q behandelt. Zur Erinnerung: IEEE 802.1D ist der Standard für MAC Access-Control-Bridges, mit denen Schicht-2-Netzwerke aufgebaut sind. Switch-Hersteller hal-

ten sich an beide Standards und fügen dann Erweiterungen, etwa zur Verwaltung, hinzu. Der IEEE 802.1Q-Standard basiert im Sprachgebrauch auf Dokumenten wie dem ISO/IEC 15802-3 -Standard für MAC-Bridges.

Bei IEEE 802.1Q wird ein 4-Byte-Header zwischen den Ethernet- und IP-Headern eingefügt. Entsprechend dem 802.1D-Standard wird er im Frame 12 Bytes tief unmittelbar hinter der Quell-MAC-Adresse eingefügt. Der Frame wird also tatsächlich verändert. Der Ethernet-Typ, der die Art der gekapselten Daten angibt, muss daher ebenfalls geändert werden. Zum Beispiel haben IP-Pakete den Ethertype-Wert 0800, doch wenn sie durch den Trunk laufen, wird er zu 8100 geändert, wie Abbildung 4-18 zeigt.



▲ **Abbildung 4-18**
Ethertype für IEEE 802.1Q

Der 802.1Q-Header ist sehr einfach und umfasst die folgenden Felder:

- den Tag-Protokoll-Identifier (TPID, 2 Byte)
- Der Wert 8100 steht direkt vor den hervorgehobenen Hexadezimalwerten.
- die Tag-Kontrollinformation (Tag Control Information, TCI, 2 Byte)

Diese Information kann auf drei Arten erzeugt werden, doch auf die von Token Ring- und FDDI-Netzwerken verwendeten gehen wir hier nicht weiter ein. Die TCI umfasst die Priorität, den kanonischen Formatindikator und die VLAN-ID. Die hexadezimale, 2 Byte lange TCI aus Abbildung 4-18 lautet *20 65*.

Priorität

Die Priorität wird bei Quality-of-Service-Implementierungen genutzt und auch Serviceklasse (Class of Service) genannt. Es handelt sich um ein Drei-Bit-Feld mit Werten von 000 (0) bis 111 (7). Voreingestellt ist die 0, auch wenn die Hersteller höhere Werte für bestimmte Arten von Traffic empfehlen. So wird beispielsweise VoIP-Traffic üblicherweise auf einen Wert von binär 101 (dezimal 5) gesetzt. Abbildung 4-18 zeigt eine leicht erhöhte Priorität von 2. Abbildung 4-19 zeigt priorisierten Traffic aus einem anderen Netzwerk. In diesem Fall ist die Priorität mit 111 (7) angegeben.

Kanonischer Formatindikator (Canonical Format Indicator, CFI)
: Dieses Ein-Bit-Feld wurde verwendet, um die mit veralteten Protokollen wie Token Ring und FDDI assoziierte Bitordnung oder die Flags für Routinginformationen anzuzeigen. Heutzutage erfolgt nahezu das gesamte Switching mittels Ethernet, weshalb dieses Feld fast nie benutzt wird und typischerweise auf 0 gesetzt ist.

VLAN-ID
: Die letzten 12 Bits sind für die VLAN-ID reserviert. Die Werte liegen zwischen 1 und 4095. Hier hat die VLAN-ID den Binärwert 1100101, was dem dezimalen VLAN 101 entspricht.

Abbildung 4-19 ▼
Getaggter Frame mit Prioritätsfeld

```
⊟ Ethernet II, Src: D-Link_b9:5c:15 (00:50:ba:b9:5c:15), Dst: D-Link_53:ff:c2 (00:50:ba:53:ff:c2)
  ⊞ Destination: D-Link_53:ff:c2 (00:50:ba:53:ff:c2)
  ⊞ Source: D-Link_b9:5c:15 (00:50:ba:b9:5c:15)
    Type: 802.1Q Virtual LAN (0x8100)
⊟ 802.1Q Virtual LAN
    111. .... .... .... = Priority: 7
    ...0 .... .... .... = CFI: 0
    .... 0000 0110 0101 = ID: 101
    Type: IP (0x0800)
⊞ Internet Protocol, Src: 192.168.16.2 (192.168.16.2), Dst: 192.168.16.1 (192.168.16.1)
⊞ Internet Control Message Protocol
```

Inter-Switch-Link (ISL)

Da ISL ein älteres, proprietäres Cisco-Protokoll ist, werden wir nicht viel Zeit auf eine Beschreibung verschwenden. Abbildung 4-20 zeigt einen mit ISL getaggten Frame und verdeutlicht diesen Tagging-Ansatz. IEEE 802.1Q nimmt ein sogenanntes »internes Tagging« vor, bei dem der VLAN-Header zwischen die Ethernet- und IP-Header eingefügt wird. Das erfordert außerdem eine Neuberechnung der Frame-CRC. ISL stellt das Tag voran. Der ISL-Header ist darüber hinaus deutlich größer als der 802.1Q-Header und kann keine Prioritäten verarbeiten. Moderne Cisco-Geräte verwenden standardmäßig IEEE 802.1Q als Trunking- und Tagging-Protokoll.

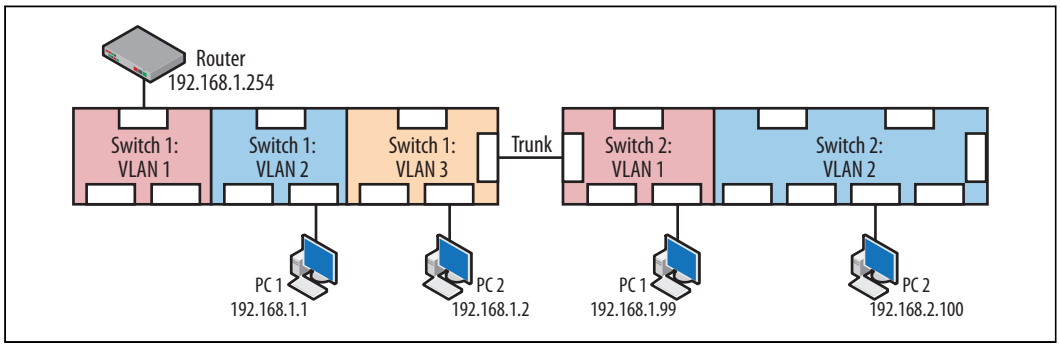
Abbildung 4-20 ▼
Mit ISL getaggter Frame

```
ISL
⊞ Destination: ISL-Frame_00 (01:00:0c:00:00:00)
  Source: Cisco_da:55:40 (00:05:32:da:55:40)
  Length: 130
  DSAP: 0xAA
  SSAP: 0xAA
  Control: 0x3
  HSA: 0x00000c
  0000 0000 0000 010. = VLAN ID: 2
  .... .... .... ...0 = BPDU/CDP/VTP: No
  Index: 0
Ethernet II, Src: Cisco_da:55:40 (00:05:32:da:55:40), Dst: Cisco_da:6c:61 (00:05:32:da:6c:61)
Internet Protocol, Src: 192.168.1.1 (192.168.1.1), Dst: 192.168.1.2 (192.168.1.2)
Internet Control Message Protocol
```

Pruning

Während ein bestimmtes VLAN durchaus über einen einzelnen Switch hinaus gehen und innerhalb der gesamten Topologie verfügbar sein kann, muss es trotzdem nicht auf jedem Switch vorhanden sein.

In Abbildung 4-21 gibt es die VLANs 1 und 2 auf beiden Switches. Doch VLAN 3(gelb) gibt es nur auf Switch 1. Den Traffic für VLAN 3 an Switch 2 weiterzuleiten ist nicht sinnvoll. Dieses sogenannte Pruning hat den Vorteil, dass der Traffic in der Trunk Line reduziert und die Sicherheit möglicherweise erhöht wird, insbesondere bei statischen Topologien. Switch 1 *beschneidet* (engl. *prune*) den Traffic von VLAN 3 an dessen Trunk-Port.



▲ **Abbildung 4-21**
Pruning-Beispiel

Die Hersteller verfolgen unterschiedliche Pruning-Ansätze. Einige erlauben standardmäßig alle VLANs (Cisco), während andere sie standardmäßig alle unterbinden. Unabhängig vom Hersteller sollten Sie aber immer die Trunking-Konfiguration untersuchen und herausfinden, welcher Ansatz für getaggte und ungetaggte Frames sowie für das Pruning der beste ist.

Erwägungen zum VLAN-Design

VLANs bauen Grenzen auf, die die Knoten oder den Datenverkehr beschneiden können. Man sollte sich daher einige Gedanken um das Design einer Multi-VLAN-Topologie machen. Die generelle Frage muss lauten: »Wer redet mit wem, und was soll damit erreicht werden?« Die folgende Liste ist als kleiner Leitfaden gedacht.

Skalierung
: Wie groß ist das Netzwerk, und wie weit muss der Traffic laufen?

Traffic-Muster
: Welche Wege nehmen die Pakete/Frames?

Anwendungen
: Warum gibt es den Traffic? Was versuchen die Hosts zu tun?

Netzwerk-Management
: Läuft SNMP oder ein anderes Management-Protokoll? Wie gelangen Sie zu allen Knoten?

Gruppen-Gemeinsamkeiten
: Was haben die Knoten gemeinsam? Gibt es gemeinsam genutzte Ressourcen oder Traffic-Muster?

IP-Adressierungsschema
: Wie sieht der IP-Adressraum aus? Wie viele Knoten liegen in jedem VLAN?

Physische Lage
: Liegen die Knoten im gleichen Büro, Gang oder Gebäude?

Statisch oder dynamisch
: Bewegen sich die Knoten, oder sind sie stationär?

Ende-zu-Ende oder lokale VLANs
: Gibt es außerhalb des Knotens liegende Knoten, die Teil eines VLANs sein müssen?

80/20- oder 20/80-Datenflussmuster
: Fließt der Großteil der Daten intern oder extern? Ändert sich dieses Muster?

Gemeinsame Sicherheitsanforderungen
: Sind die Knoten Server? Endgeräte? Wireless-Geräte? Stellen die Knoten wichtige Unternehmensressourcen dar? Sind sie öffentlich zugänglich?

Quality of Service
: Gibt es irgendwelche Erwägungen in Bezug auf die Servicequalität?

Neben diesen allgemeinen Fragen gibt es weitere Vorgehensweisen, die Sicherheitsrisiken mindern und wichtige Netzwerk-Ressourcen schützen:

- Wireless sollte in einem eigenen VLAN liegen. Da Wireless ein gemeinsam genutztes Medium (Shared Media) ist, werden auch der gesamte Broadcast- und ein Großteil des Multicast-Traffics geteilt. Darüber hinaus sehen alle Wireless-Knoten jeglichen »gefluteten« Unicast-Traffic. Der Aufbau eines VLANs für die Wireless-Knoten schränkt den Traffic ein, den diese Geräte

sehen können. Darüber hinaus müssen mögliche Angriffe auf den Wireless-Bereich erst einmal die gesetzte Grenze überwinden, bevor sie andere Teile des Netzwerks erreichen.

- VoIP-Elemente sollten ebenfalls in einem eigenen VLAN liegen. Das dient aber eher der Servicequalität denn der Sicherheit. Wann immer Sprachdaten um Bandbreite konkurrieren, besteht die Gefahr, dass die Performance einbricht. Auch Sicherheitsaspekte werden durch das VLAN zum Teil entschärft. Tools wie Wireshark können Sprachdaten nicht nur festhalten, sondern auch dekodieren und abspielen. Es ist daher wichtig, Sprachdaten nach Möglichkeit getrennt vorzuhalten.
- Andere wichtige Netzwerk-Geräte wie Server, oder auch Nutzer mit sensitiven Daten, sollten ebenfalls in eigenen VLANs liegen. Neben den bereits erwähnten Gründen bieten viele Hersteller außerdem Features an, die den Aufbau von VLANs mit bestimmten Sicherheits- und QoS-Policies erlauben.

Sicherheitserwägungen

Dieses Kapitel hat die Notwendigkeit der Isolation von Traffic behandelt. Unternehmen müssen nicht alle Daten an jeden Port weiterleiten. Das ist ineffizient und stellt ein potenzielles Sicherheitsrisiko (durch Lauscher) dar. Es gibt verschiedene Konfigurationspunkte, die bei keiner Checkliste für den VLAN-Einsatz fehlen dürfen. Eine der größten Herausforderungen beim Einsatz von Netzwerk-Geräten besteht darin, ihr Standardverhalten zu verstehen. Switches und Router sind da keine Ausnahme, insbesondere durch die steigende Zahl von Features.

Einer dieser Punkte ist der Standard-Konfigurationsmodus der Ports eines Switches. Die meisten Switch-Ports werden mit Computern verbunden und fungieren daher als *Access*-Ports. Nicht ganz so offensichtlich ist aber, dass bei vielen Geräten die Standardkonfiguration nicht *Access*, sondern *dynamisch* lautet. Das bedeutet, dass der Port bereit ist, den Betriebsmodus auszuhandeln. Sind zwei Switches miteinander verbunden und ist einer dieser Switches mit einem Trunk-Port konfiguriert, dann werden häufig dynamische Trunking-Protokoll-Nachrichten generiert. Sobald diese Nachricht einmal empfangen wurde, kann es sein, dass der zweite Switch seinen Port *automatisch* in einen Trunk-Port umwandelt. Das ist in Abbildung 4-22 zu sehen.

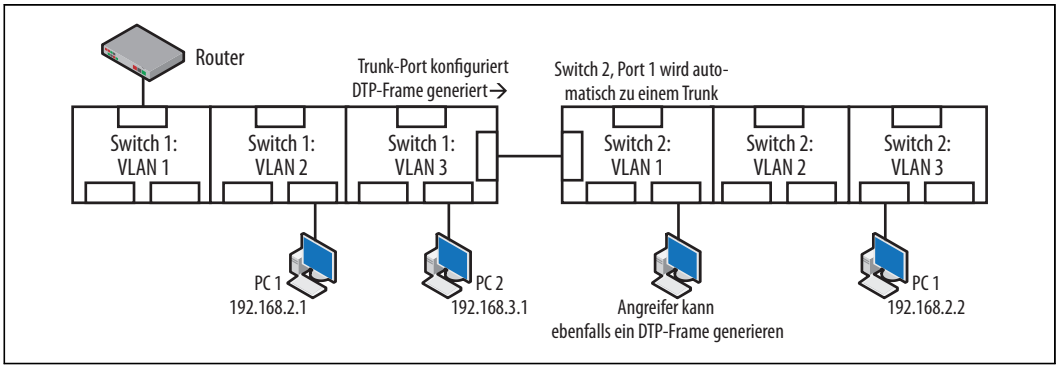


Abbildung 4-22 ▲
Sicherheitsrisiko durch dynamische Port-Konfiguration

Diese Autokonfiguration klingt zuerst einmal praktisch, doch was sollte einen Angreifer daran hindern, die gleiche Nachricht zu senden und einen Port auf die gleiche Weise umzustellen? Der Port des Angreifers empfängt dann Broadcast-, Multicast- und gefluteten Unicast-Traffic für alle nicht dem Pruning unterliegenden VLANs. Ein Angreifer kann auf diese Weise nicht nur mehr über Ihr Netzwerk erfahren, sondern kann auch getaggte Frames erzeugen, die über das gesamte Netzwerk ausgeliefert werden können. Die dynamische Konfiguration sollte wann immer möglich ausgeschaltet werden.

Neben dem Pruning an den entsprechenden VLAN-Grenzen und der Standard-Konfiguration der Ports kann es klug sein, zusätzliche Konfigurationsänderungen vorzunehmen. Ungenutzte Ports können in einem »toten VLAN« gesammelt werden, das nicht geroutet wird und vom Netzwerk abgeschnitten ist. Viele Hersteller bieten zusätzliche Sicherheitserweiterungen für die Ports an, etwa autorisierte MAC-Adressen oder eine Beschränkung der erlaubten MAC-Adressen. Tauchen an einem Port ungültige MAC-Adressen auf, wird er automatisch heruntergefahren oder deaktiviert.

Lektüre

Der IEEE 802.1Q-Standard heißt eigentlich »IEEE Standards for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks«.
ISO/IEC 15802-3 ANSI/IEEE Std 802.1D: »Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Common specifications – Part 3: Media Access Control (MAC) Bridges«

Zusammenfassung

VLANs sind ein grundlegendes Werkzeug zum Aufbau von Netzwerkgrenzen. Zwar können sie Sie in Bezug auf die Weiterleitung von Traffic vor große Herausforderungen stellen, sie sind aber gleichzeitig in Sachen Sicherheit und Servicequalität ein mächtiges Werkzeug. Dieses Kapitel hat den Betrieb und die Methoden der Propagation von VLANs durch große Topologien behandelt. Beim Einsatz von VLANs und Trunks sind verschiedene Design-Erwägungen zu beachten. Die grundlegende Frage lautet: »Wer redet mit wem und warum?« Mit den Topologien und den VLANs wächst auch die Komplexität. Es ist wichtig, das Standardverhalten und die Standardkonfiguration der Netzwerk-Elemente im Auge zu behalten, damit lokale Konfigurationen das Netzwerk nicht einer Gefahr aussetzen.

Fragen

1. Broadcast-Frames werden weiterpropagiert, bis sie ein geroutetes Interface erreichen.
   a. WAHR
   b. FALSCH
2. Broadcast- und Multicast-Traffic läuft über VLAN-Grenzen hinweg, Unicast-Traffic hingegen nicht.
   a. WAHR
   b. FALSCH
3. Standardmäßig sind alle Hosts mit dem gleichen VLAN verbunden.
   a. WAHR
   b. FALSCH
4. Hosts wissen üblicherweise nicht, mit welchem VLAN sie verbunden sind.
   a. WAHR
   b. FALSCH
5. In einem modernen Datennetzwerk besteht die primäre Aufgabe einer Trunk Line in der Übertragung von VLAN-Informationen.
   a. WAHR
   b. FALSCH

6. Zwar sind SATs und VLANs beide Teil eines Switches, haben aber nichts miteinander zu tun.
    a. WAHR
    b. FALSCH
7. Welches ist das Industriestandard-Trunking-Protokoll?
    a. ISL
    b. IEEE 802.1
    c. VLAN
8. Pruning ist eine Technik, die den nicht autorisierten Zugriff auf Trunk-Lines unterbindet.
    a. WAHR
    b. FALSCH
9. Der dynamische Portmodus stellt ein Sicherheitsrisiko dar, da Angreifer den gesamten VLAN-Traffic sehen können, der nicht dem Pruning unterliegt.
    a. WAHR
    b. FALSCH
10. Drahtlos- und VoIP-Dienste sollten in eigene VLANs gelegt werden.
    a. WAHR
    b. FALSCH

Antworten

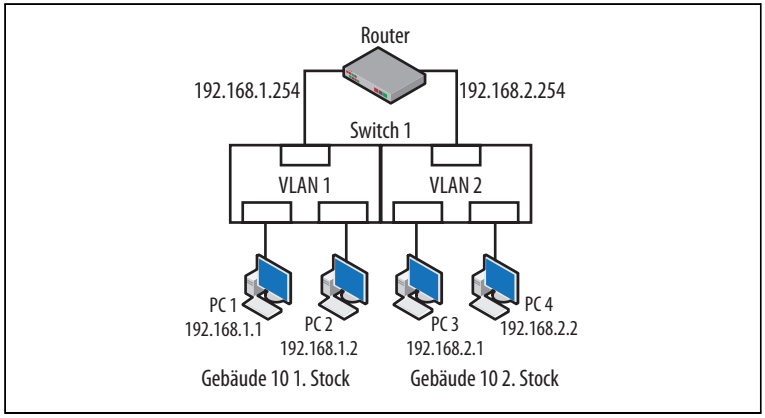
1. WAHR
2. FALSCH
3. WAHR
4. WAHR
5. WAHR
6. FALSCH
7. IEEE 802.1
8. FALSCH
9. FALSCH
10. WAHR

Laborübungen

Übung 1: Ein lokales VLAN einrichten

Material: Ein VLAN-fähiger Switch und ein Router. Ein Heim-Gateway kann verwendet werden, wenn man es in einen Router umwandeln kann, um mit dem NAT-Betrieb nicht durcheinanderzukommen.

Hinweis: Das Ziel dieser Übung besteht einfach darin, die Grundkonfiguration zu verstehen, die für ein Routing zwischen VLANs ohne Trunks notwendig ist (siehe Abbildung 4-23).



◀ **Abbildung 4-23**
Übung 1

1. Legen Sie auf dem Switch zwei VLANs an.
2. Fügen Sie einen Host in jedes VLAN ein, und ermitteln Sie das IP-Adressierungsschema. So könnte ein VLAN beispielsweise 192.168.1.0 verwenden und das andere 192.168.2.0. Praktischer Cisco-Befehl: `switchport access vlan X`.
3. Verbinden Sie eine Router-Schnittstelle mit jedem VLAN, und weisen Sie ihr die richtigen IP-Adressen zu. An diesem Punkt sollten sich die Knoten der verschiedenen Netzwerke gegenseitig »anpingen« können.

Übung 2: VLANs und die SAT

Material: Ein VLAN-fähiger Switch und ein Router.

1. Sobald die Topologie aus Übung 1 steht, lassen sich alle Knoten und Router-Schnittstellen anpingen.
2. Untersuchen Sie die SAT des Switches. Praktischer Cisco-Befehl: `show mac-address-table`

3. Vergleichen Sie die Tabelle mit einer, bei der alle Knoten im gleichen VLAN liegen.
4. Entwickeln Sie mit den Informationen in der SAT und der Routingtabelle des Routers eine Schritt-für-Schritt-Prozedur zum Forwarding von Paketen von einem Computer zum anderen.

Übung 3: Was sehen Sie?

Material: Ein VLAN-fähiger Switch, ein Router und Wireshark.

Das Ziel dieser Übung besteht darin, herauszufinden, wie weit der Traffic in einem VLAN reist und ob er in einem anderen VLAN am gleichen Switch zu sehen ist.

1. Starten Sie ein Capture auf einem der Netzwerk-Hosts in einem der VLANs.
2. Im anderen VLAN erzeugen Sie Broadcast-Traffic, indem Sie eine ungenutzte IP-Adresse im gleichen Netzwerk anpingen, was einen ARP-Request generiert.
3. Auf dem gleichen Host generieren Sie Unicast-Traffic, indem Sie den Router anpingen.
4. Wie sich herausstellt, generieren Windows-PCs periodisch Multicast-Traffic, wenn sie nach Diensten suchen.
5. Hat der Capture-Knoten im anderen VLAN den Unicast-, Multicast- oder Broadcast-Traffic des Quell-Hosts gesehen? Die Antwort muss »NEIN« lauten.
6. Als zusätzliches Experiment ändern Sie die IP-Adresse des Capture-Hosts so ab, dass er im gleichen Netzwerk wie der Quell-Host liegt. Die Rechner liegen nun also im gleichen IP-Netzwerk, aber in verschiedenen VLANs. Versuchen Sie nun, die beiden Knoten anzupingen. Dieser Versuch sollte scheitern, da die Rechner zwar im gleichen Netzwerk liegen, aber durch den Switch getrennt wurden und Traffic die VLAN-Grenze nicht überschreiten kann.

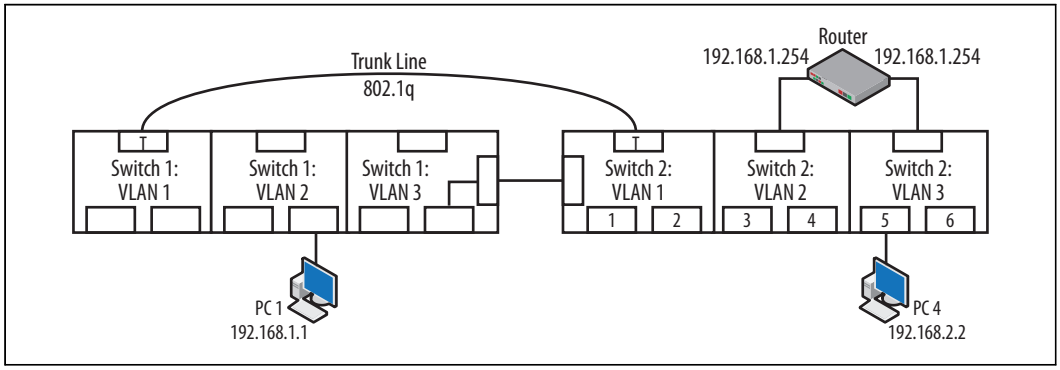
Übung 4: Einfaches Trunking

Material: Ein zweiter VLAN-fähiger Switch, ein Trunk-fähiger Switch und ein Router.

1. Verbinden Sie einen weiteren Switch mit der bereits bestehenden Topologie.
2. Auf dem neuen Switch legen Sie die gleichen VLANs an.

3. Binden Sie einen Host in jedes VLAN ein. Wenn Sie nicht genug Computer haben, reicht es aus, einen in einem VLAN am ersten Switch und den zweiten in dem anderen VLAN am neuen Switch anzuschließen (siehe Abbildung 4-24).

▼ **Abbildung 4-24**
Übung 4



4. An beiden Switches konfigurieren Sie die Ports als Trunk-Ports, die die Switches miteinander verbinden. Praktischer Cisco-Befehl: `switchport mode trunk`, `switchport trunk encapsulation dot1q`
5. An diesem Punkt sollten die Netzwerk-Hosts in der Lage sein, sich gegenseitig anzupingen.
6. Als zusätzliche Übung können Sie die Fähigkeiten der Switches untersuchen. Versuchen Sie, einen Host einzurichten, der den über den Trunk laufenden Traffic festhalten kann. Das geschieht üblicherweise mit einem Mirror- oder Monitor-Port. Das Ziel ist, die im Trunk verwendeten IEEE 802.1Q-Tags zu untersuchen. Praktischer Cisco-Befehl: `monitor session`

KAPITEL 5

Routing Information Protocol

In diesem Kapitel:

- Version 1 versus Version 2
- Protokoll-Beschreibung
- Struktur
- Grundlegender Betrieb
- Fortgeschrittener Betrieb
- Wie komme ich aus einem Netzwerk raus?
- RIP und Schleifen
- Sicherheit
- RIP und IPv6
- Lektüre
- Zusammenfassung
- Fragen
- Antworten
- Laborübungen

Um die beste Route definieren zu können, brauchen wir natürlich eine Möglichkeit, die Güte messen zu können.

RFC 1058

Das Routing Information Protocol, kurz RIP, ist ein internes Distanzvektor-Protokoll für kleine Netzwerke. Es ist in den IETF-RFCs 1058, 1388 und 1723 definiert. Es war eines der ersten im Internet verwendeten Routingprotokolle. Das Protokoll hat zwei Versionen durchlaufen, um auch klassenfreie (classless) Adressräume verarbeiten zu können. Dieses Kapitel behandelt die Konstruktion, den Betrieb und den Inhalt (über Paket-Captures) des Protokolls. Die RFCs zur Version 2 von RIP gibt es seit ca. 1998. Selbst zu dieser Zeit wurde behauptet, RIP sei ein unterlegenes Protokoll, das seinen Zenit bereits überschritten habe. Dennoch hatte RIP immer noch seine Fans. Hier ein (frei übersetztes) Zitat aus RFC 2453:

> Seit der Einführung von OSPF und IS-IS gibt es Leute, die RIP für veraltet halten. Zwar stimmt es, dass die neueren IGP-Routingprotokolle RIP weit überlegen sind, aber RIP hat auch einige Vorteile. Insbesondere bei kleinen Netzwerken hat RIP nur sehr wenig Overhead in Bezug auf Bandbreite sowie Konfigurations- und Verwaltungsaufwand. RIP ist sehr einfach zu implementieren, insbesondere im Vergleich zu den neueren IGPs.
>
> Darüber hinaus sind sehr viel mehr RIP-Implementierungen im Einsatz als bei OSPF und IS-IS zusammen. Es ist sehr wahrscheinlich, dass das noch einige Jahre so bleibt. Wenn wir davon ausgehen, dass RIP für viele Umgebungen noch eine Zeit lang nützlich sein wird, ist es vernünftig, auch die Brauchbarkeit von RIP zu erhöhen. Das gilt umso mehr, als der Vorteil weitaus höher ist als die Kosten der Änderungen.

Und das war vor der Implementierung von RIPv2. RIP ist seitdem in anderen Standards wie dem *High Assurance Internet Protocol En-*

cryptor Interoperability Standard oder HAIPE IS aufgegangen. Zusätzlich wurde mit den RFCs 2082 und 4822 die Sicherheit von RIPv2 verbessert. Diese Bemühungen weisen darauf hin, dass uns RIPv2 noch einige Zeit erhalten bleibt. Aber auch wenn RIP nicht die Welt beherrscht, ist es eine gute Quelle und Lernumgebung für das Routing.

Version 1 versus Version 2

RIP gibt es schon sehr lange. Trotz seines Erfolges war es nicht ohne Schwächen, und die RIP-Version 1 wurde durch die RIP-Version 2 ersetzt. RFC 1923 behandelt die Eignung bzw. Nicht-Eignung von RIPv1. Die Probleme mit RIPv1 leiten sich alle aus der Klassenorientierung oder dem strikten Festhalten an Class A-, B- und C-Netzwerken ab. RIPv1-Nachrichten enthalten keine Netzwerkmasken, und somit fehlt ihnen die Flexibilität moderner Ansätze zur Verwaltung von Adressraum. RFC 1923, RIPv1, lässt sich wie folgt zusammenfassen:

- RIPv1 nimmt an, dass die lokal verwendete Maske auch die Maske der ganzen Gruppe von Netzwerken ist.
- RIPv1 kann nicht mit Subnetzen variabler Länge, beim Supernetting und beim CIDR (Classless Inter-Domain Routing) verwendet werden.

Darüber hinaus wird RIPv1 als »einfaches Distanzvektor-Protokoll« bezeichnet, was bedeutet, dass es selbst mit Erweiterungen wie Split Horizon und Poison Reverse zeitraubende Techniken wie Count to Infinity nutzen muss, um zu konvergieren. Die Schlussfolgerung des RFCs lautet, dass man, wenn man ein Distanzvektor-Protokoll nutzen muss, mit RIPv2 arbeiten und dabei nur die einfachsten Sicherheitsfeatures aktivieren soll. Dieses Kapitel schaut sich beide Versionen in Form von Paketen an, da RIPv1 die Standardeinstellung ist. Die klare Empfehlung lautet aber, mit RIPv2 zu arbeiten. Die Ideen, die hinter Split Horizon, Poison Reverse und Count to Infinity stehen, werden später in diesem Kapitel behandelt.

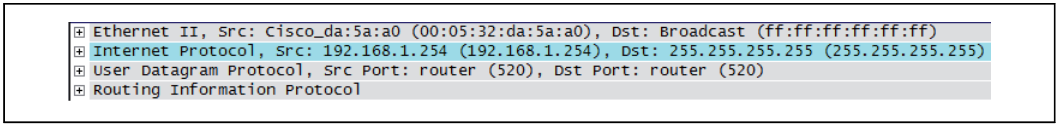
Protokoll-Beschreibung

Die Geschichte von RIP beginnt üblicherweise mit RFC 1058, doch das RFC versucht tatsächlich nur Ideen zu konsolidieren, die bereits verwendet wurden. Eine dieser Ideen (das Distanzvektoren nutzende »routed« von Berkeley Unix) war zu jener Zeit der De-facto-Standard für das Routing. Doch selbst 1988 ging man davon aus,

dass RIP für das Routing über große Netze nicht geeignet sein würde. Der Ansatz sollte eher so aussehen, dass ein autonomes System (AS) ein internes Gateway-Protokoll (Interior Gateway Protocol, IGP) wie RIP sowie ein anderes Routingprotokoll für die Kommunikation mit anderen AS-Netzwerken verwendet. Hier ein (frei übersetztes) Zitat aus RFC 1058:

> RIP wurde für den Einsatz in moderat dimensionierten Netzwerken mit halbwegs homogener Technik entwickelt. Es eignet sich daher als IGP für viele (Firmen-)Standorte oder regionale Netzwerke, die mit seriellen Leitungen arbeiten, deren Geschwindigkeit nicht groß variiert.

RIP ist ein Distanzvektor-Protokoll. Distanzvektor-Protokolle implementieren üblicherweise den Bellman-Ford-Algorithmus, um die besten Pfade zu ermitteln. Doch die Klasse von Protokollen wurde bereits vorher in *Flow in Networks* von Ford und Fulkerson definiert. Auch wenn sich die Abstammungslinie bis zu Xerox-Netzwerken zurückdatieren lässt, wurde RIP für das IP-Routing entworfen. RIP ist ein Routingprotokoll, das Tabellen austauscht, um benachbarte Router zu aktualisieren. Jeder Router sendet dabei seine eigene Routingtabelle per UDP (User Datagramm Protocol) über seine aktiven Schnittstellen hinaus. Abbildung 5-1 zeigt die verwendete Kapselung.



▲ **Abbildung 5-1**
RIP-Kapselung

Router, die diese Informationen empfangen, entscheiden selbst, ob sie ihre eigenen Tabellen aktualisieren oder nicht. Router verwenden die Quell-IP-Adresse des IP-Pakets als den weiterleitenden Router. Erinnern Sie sich aus Kapitel 1 daran zurück, dass die IP-Adressen der weiterleitenden Router für den nächsten Hop von entscheidender Bedeutung sind. Informationen, die die Präfixlänge oder die Metrik verbessern, werden gespeichert. Das setzt voraus, dass die administrative Distanz im »gesamten RIP-Netzwerk« gleich ist. Die neue Netzwerk-Information kann Teil zukünftiger Updates sein. Der einfache Austausch von Routingtabellen kann ebenso viele Probleme verursachen, wie der Wechsel zum dynamischen Routing behebt. Aus diesem Grund kennt RIP auch verschiedene Mechanismen zur Konvergenzbeschleunigung und zur Vermeidung von Schleifen. Dazu zählen auch die weiter oben erwähnten Techniken wie Split Horizon, Poisoning und Count to Infinity.

RIP-Netze sind auf eine Größe von 15 Hops beschränkt. Das bedeutet, soweit es RIP betrifft, dass 16 für *Unendlich* oder *Unerreich-*

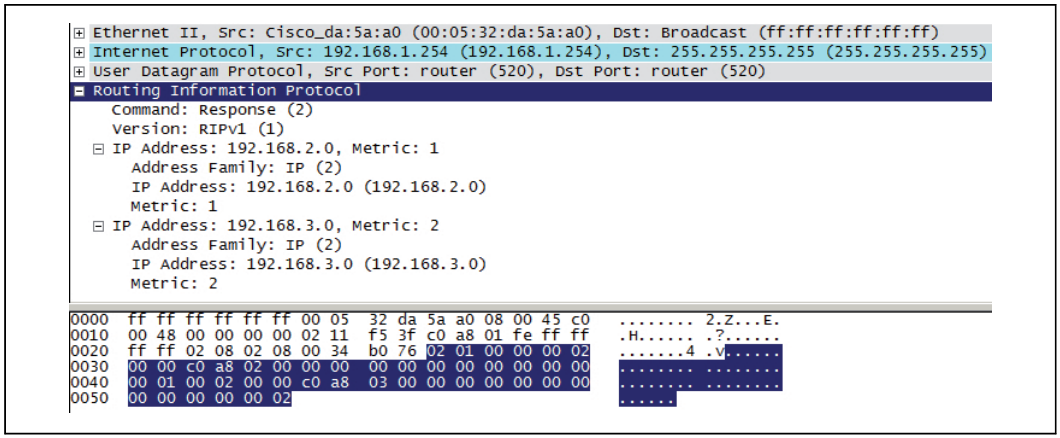
bar steht. Jedem durchlaufenen Netzwerk wird der Wert von einem Hop zugeordnet. Dieser Hop-Zähler ist die »Metrik«, die von RIP zur Messung der Distanz verwendet wird. RIP berücksichtigt keinerlei Echtzeit-Daten wie Kosten, Auslastung oder Geschwindigkeit. Daher wird jeder Weg nach dem gleichen Standard gemessen. Router empfangen RIP-Updates von direkt verbundenen benachbarten Routern. Sobald ein Router ein Update empfängt, sendet er sein eigenes Update. Bevor ein Router sein eigenes aktualisiertes Routing anbieten kann, muss er die Metrik aller gelernten Routen um 1 erhöhen. Das neue Update wird mit der IP-Adresse des neuen Routers gesendet. Diese IP-Adresse wird zum »Nächster Hop«-Router, der in die Routingtabelle des benachbarten Routers eingetragen wird, und die Metrik entspricht der Distanz zum Ziel über diese IP-Adresse.

Denken Sie daran, dass ein Routingtabelleneintrag Informationen über das Alter der Information, die Zieladresse, den nächsten Hop oder das nächste Gateway aus Sicht des Routers, die zum Erreichen des nächsten Hops verwendete lokale Schnittstelle und die Kosten der Route enthält. Mithilfe dieser Information kann der Router eine »Distanzvektor-Entscheidung« zur Brauchbarkeit einer Route treffen. Da diese Information an benachbarte Router gesendet wird und alle Updates ebenfalls gesendet werden, kann man die gesamte Menge der Netzwerke verstehen, indem man nur mit benachbarten Routern spricht.

Die RIP zugeordnete administrative Distanz (der Protokoll-Wert) ist 120. Diese Information erscheint in der Routingtabelle zusammen mit den Präfixlängen und den Metriken.

Struktur

Wie Sie in Abbildung 5-2 sehen, haben RIPv1-Pakete eine einfache Struktur. Das abgebildete Paket wurde in einer frühen Konfigurationsphase der in diesem Kapitel verwendeten Topologie festgehalten. An diesem Punkt war das Netzwerk nur mit RIP Version 1 konfiguriert.



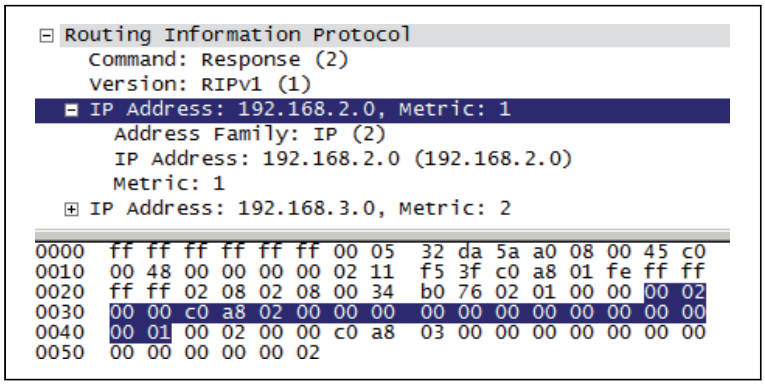
▲ **Abbildung 5-2**
RIPv1-Paket

Befehl (Command)
: Ein 1-Byte-Feld, das den Nachrichtentyp beschreibt. Ein Request fordert eine Routingtabelle an und eine Response enthält die Routingtabelle des Routers. Es sind noch eine Reihe anderer Nachrichten definiert, doch diese sind mittlerweile veraltet.

Version
: Dieses ebenfalls ein Byte lange Feld gibt die verwendete RIP-Variante an.

Null (Zero)
: Auf das Versionsfeld und die Adressfamilien-ID folgen »Muss-Null-sein-Felder«. Diese sind 2 Byte lang. Ein 8 Byte langes »Muss-Null-sein-Feld« folgt auf die tatsächliche IP-Adresse des Zielnetzwerks.

Jeder Eintrag in der Routingtabelle hat Platz für Informationen zum Netzwerk und zu dessen Metrik. Die Hexwerte für das 192.168.2.0-Netzwerk sind in Abbildung 5-3 zu sehen.



◀ **Abbildung 5-3**
Hex-Beispiel für Netzwerk 192.168.2.0

Adressfamilien-ID (AFI)
: Dieser Wert gibt den Typ des im aktuellen Netzwerk verwendeten Kommunikationsprotokolls an. Obwohl Platz dafür wäre, andere Protokolle aufzuführen, wurden in RFC 1058 keine anderen definiert. Die AFI für IP ist 2.

IP-Adresse
: Die IP-Adresse des Zielnetzwerks in der Routingtabelle. Im obigen Hex-Beispiel hat das 192.168.2.0-Netzwerk den Wert c0 a8 02 00.

Metrik
: Die Distanz zum Zielnetzwerk in Hops. Im Beispiel liegt die Anzahl der Hops bei 1. Dies ist ein 4-Byte-Feld.

RIPv1-Pakete sind auf eine Gesamtlänge von 512 Bytes beschränkt. Bei großen Routingtabellen werden die Einträge auf mehrere Pakete aufgeteilt.

Die Struktur des RIPv2-Pakets in Abbildung 5-4 ist ähnlich, enthält zusätzlich aber noch eine Reihe von Feldern für Subnetze. Der Konsistenz halber zeigt dieses Paket die gleiche Netzwerkadresse.

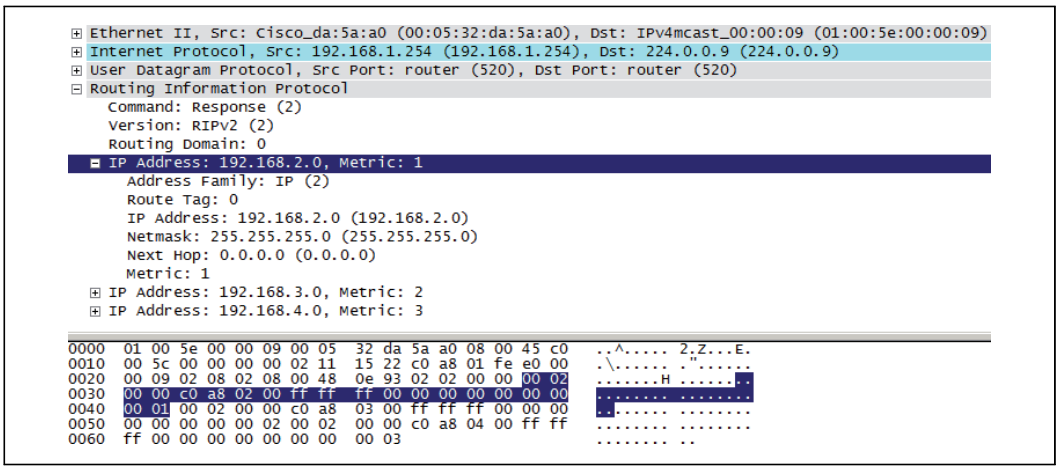


Abbildung 5-4 ▲
RIPv2-Paket

Das Nachrichtenformat der beiden Versionen ist grundsätzlich gleich, d.h., die in RFC 1058 definierten Felder bleiben unverändert bestehen. Vergleicht man den Hex-Teil der Pakete aus Abbildung 5-3 und Abbildung 5-4, dann sieht man, dass bei beiden Versionen die gleiche Anzahl von Bytes zugewiesen wurde. Die Änderungen des globalen Pakets bei RIPv2 umfassen den Versionswert und das Routingdomänen-Feld.

Routingdomäne
: Neben dem Routen-Tag für einzelne Ziele differenziert die RIP-Routingdomäne zwischen der aktuellen Gruppe von RIP-Netzwerken und denjenigen, die über externe Protokolle gelernt wurden.

Für individuelle Netzwerke wurden Felder für die Netzmaske, die Route und den nächsten Hop hinzugefügt.

Netzmaske
: Die Maske des Zielnetzwerks. Es herrscht eine gewisse Besorgnis, dass dieses Feld von RIPv1-Routern fehlinterpretiert werden könnte, weshalb in einer gemischten Umgebung eine gewisse Vorsicht angezeigt ist. Oder verwenden Sie einfach nur RIPv2.

Routen-Tag
: Das Routen-Tag-Feld ist ein Attribut, das eine Route identifiziert, die über eine externe Quelle (etwa einen anderen IGP) gelernt wurde. Die Route stammt nicht aus der aktuellen Gruppe von RIP-Netzwerken.

Nächster Hop
: Normalerweise verwendet ein Router, der eine RIP-Nachricht empfängt, die Quell-IP-Adresse als nächsten Hop für Routingtabellen-Einträge. Hat dieses Feld den Wert 0.0.0.0, verwendet der Router die Quell-IP-Adresse des Updates als nächsten Hop. Es gibt Zeiten, in denen es mehr als einen Weg zum Ziel gibt, und in diesem Fall müssen die Quell-IP-Adresse und der nächste Hop nicht übereinstimmen. Die Nächster-Hop-Adresse muss aber auf jeden Fall aus dem Netzwerk erreichbar sein, dem es bekanntgegeben wird.

Ein abschließender Hinweis zur Adressfamilien-ID für RIPv2: RIPv2 erlaubt die Authentifizierung von RIPv2-Nachrichten. Wird die AFI auf den Wert FFFF gesetzt, dann wird der für das Netzwerkziel bereitgestellte Platz (20 Bytes) für die Authentifizierungsinformationen genutzt. Sie umfassen einen 2-Byte-Authentifizierungstyp und 16 Bytes mit Authentifizierungsdaten.

Grundlegender Betrieb

Wie vorhin beschrieben wurde, verwendet RIP einen Tabellenaustausch, um seine Nachbarn mit Updates zu erreichbaren Netzwerken zu versorgen. Die in Abbildung 5-5 abgebildete Topologie wird verwendet, um den grundlegenden Betrieb von RIP durchzugehen und einige Techniken vorzustellen, die RIP in Bezug auf die Per-

formance optimieren. Da RIPv1 nicht mehr verwendet werden sollte, nutzen alle diskutierten Beispiele RIPv2. Die Topologie besteht aus vier Netzwerken. Die IP-Adressen der Router-Schnittstellen sind ebenfalls aufgeführt. Sie kennen sie vielleicht noch aus Kapitel 1, und unsere Diskussion beginnt auf die gleiche Weise.

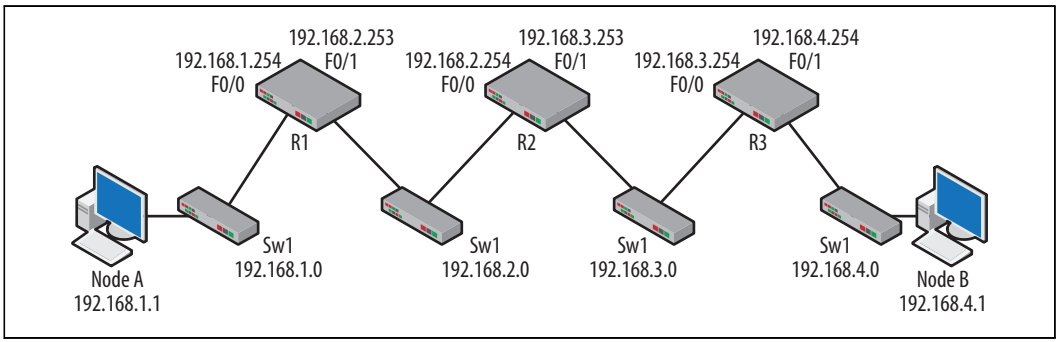


Abbildung 5-5 ▲
RIP-Topologie

Am Anfang werden die Router mit ihren IP-Adressen konfiguriert, aber RIP läuft noch nicht. Die Routingtabellen der Router enthalten nur die direkt verbundenen Routen. Jeder Router ist sich der beiden Netzwerke bewusst, für die er Schnittstellen besitzt. Nebenbei bemerkt erscheint der Begriff »direkt verbundene Router« in einem frühen RFC und ist keine Cisco-Sache.

Tabelle 5-1: Anfängliche Routingtabellen

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1

Von links nach rechts gehend, wird nun RIPv2 auf den Routern konfiguriert. Die Cisco-Befehle sind einfach und würden im Falle von R1 wie folgt aussehen:

```
router rip
    version 2
    network 192.168.1.0
    network 192.168.2.0
```

Sobald diese Befehle abgesetzt wurden, werden RIP-Pakete über beide Schnittstellen von R1 gesendet. Doch obwohl R2 diese Pakete sieht, wird seine Routingtabelle noch nicht aktualisiert, weil kein RIP läuft.

Tipp

Aktuelle Versionen von Cisco-IOS kennen einen `auto-summary`-Befehl für RIP. Dieser Befehl ist standardmäßig aktiv und »fasst Subpräfixe der klassenorientierten Netzwerkgrenzen zusammen,

wenn klassenorientierte Netzwerkgrenzen überschritten werden«. Beim Routing zwischen nicht miteinander verbundenen Subnetzen muss dieser Befehl deaktiviert werden, damit die Subnetze berücksichtigt werden.

Die von R1 generierten Pakete haben eine bestimmte Reihenfolge und gehorchen (wie wir noch sehen werden) der Split Horizon-Regel. Die ersten Pakete sind in Abbildung 5-6 zu sehen und wurden im 192.168.1.0-Netzwerk festgehalten.

12	192.168.1.254	224.0.0.9	RIPv2	9.996674	Request
57	192.168.1.254	224.0.0.9	RIPv2	17.821149	Response
58	192.168.1.254	224.0.0.9	RIPv2	18.041530	Response
71	192.168.1.254	224.0.0.9	RIPv2	47.610498	Response
78	192.168.1.254	224.0.0.9	RIPv2	75.023898	Response
83	192.168.1.254	224.0.0.9	RIPv2	102.317097	Response
89	192.168.1.254	224.0.0.9	RIPv2	129.021237	Response
95	192.168.1.254	224.0.0.9	RIPv2	154.547572	Response
102	192.168.1.254	224.0.0.9	RIPv2	180.903084	Response

▲ **Abbildung 5-6**
RIPv2-Austausch beim Start

Bei dieser Ausgabe wurden die RIP-Pakete herausgefiltert, weshalb es so aussieht, als wären einige Pakete ausgelassen worden. Das erste Paket sendet einen Request. Der Nachrichtentyp fordert den benachbarten Router auf, ihm seine Routingtabelle bereitzustellen. Alle Pakete stammen vor R1, d.h., es ist keine Response eingegangen. Sobald R1 ein Netzwerk zugewiesen wurde, generiert es eine Response, die die Routingtabelle von R1 enthält. Diese Nachrichten sind in Abbildung 5-7 und Abbildung 5-8 zu sehen.

▼ **Abbildung 5-7**
RIP-Request

```
Ethernet II, Src: Cisco_da:5a:a0 (00:05:32:da:5a:a0), Dst: IPv4mcast_00:00:09 (01:00:5e:00:00:09)
Internet Protocol, Src: 192.168.1.254 (192.168.1.254), Dst: 224.0.0.9 (224.0.0.9)
User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
Routing Information Protocol
  Command: Request (1)
  Version: RIPv2 (2)
  Routing Domain: 0
⊞ Address not specified, Metric: 16
```

```
Ethernet II, Src: Cisco_da:5a:a0 (00:05:32:da:5a:a0), Dst: IPv4mcast_00:00:09 (01:00:5e:00:00:09)
Internet Protocol, Src: 192.168.1.254 (192.168.1.254), Dst: 224.0.0.9 (224.0.0.9)
User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
Routing Information Protocol
  Command: Response (2)
  Version: RIPv2 (2)
  Routing Domain: 0
⊞ IP Address: 192.168.2.0, Metric: 1
```

▲ **Abbildung 5-8**
RIP-Response

Request-Nachrichten können die gesamte oder einen Teil der Routingtabelle anfordern und werden Eintrag für Eintrag verarbeitet. Gibt es nur einen Zieleintrag mit der AFI 0 und der Metrik 16, dann wird die gesamte Routingtabelle angefordert. Response-Nachrich-

ten werden immer gesendet, sobald ein Request eingeht, auch während eines Updates und im normalen Betrieb.

Beim Empfang einer Response-Nachricht muss der Router den Inhalt validieren, da diese Information in die Routingtabelle eingehen kann. Beispielsweise könnte die Quell-IP-Adresse und deren Format untersucht werden. An diesem Punkt werden die Metriken und Präfixlängen geprüft. Gibt es keine vergleichbaren Einträge oder sind die Response-Werte besser, dann werden diese Routen installiert. Die Timer werden akualisiert (siehe unten), und ein Update wird gesendet, nachdem die Metriken erhöht wurden.

Sobald R2 und R3 mit einem ähnlichen Satz von Befehlen konfiguriert wurden (die Netzwerke sind andere), werden die Routingtabellen anhand der empfangenen Informationen aktualisiert. Zusätzlich werden ähnliche Pakete zwischen den Routern generiert. Es gibt eine Abweichung von dem bisher gesehenen Traffic: Sobald die Router ihre Nachbarn kennen, die ebenfalls RIPv2 ausführen, werden die Nachrichten direkt an diese benachbarten Router geschickt (siehe Abbildung 5-9).

No.	Time	Source	Destination	Protocol	Info
8	24.819831	192.168.2.253	224.0.0.9	RIPv2	Request
9	26.645505	192.168.2.253	224.0.0.9	RIPv2	Response
19	56.615399	192.168.2.253	224.0.0.9	RIPv2	Response
28	82.342309	192.168.2.253	224.0.0.9	RIPv2	Response
36	107.876897	192.168.2.253	224.0.0.9	RIPv2	Response
40	111.708452	192.168.2.254	224.0.0.9	RIPv2	Request
41	111.710017	192.168.2.253	192.168.2.254	RIPv2	Response
48	126.018395	192.168.2.254	224.0.0.9	RIPv2	Response
51	136.785207	192.168.2.253	224.0.0.9	RIPv2	Response
57	150.014075	192.168.2.254	224.0.0.9	RIPv2	Response
63	166.078039	192.168.2.253	224.0.0.9	RIPv2	Response
66	178.717430	192.168.2.254	224.0.0.9	RIPv2	Response

Abbildung 5-9 ▲
Paketaustausch zwischen R2 und R3

Dieser Paket-Satz beginnt am Anfang unserer Konfiguration mit dem ersten gesendeten Request (Paket 8), nachdem R1 für RIP konfiguriert wurde. Beachten Sie die Quell-IP-Adresse dieses Pakets. Paket 40 wurde gesendet, nachdem R2 für RIP konfiguriert wurde. Das daraus resultierende Response-Paket (41) ist nicht an die RIPv2-Multicast-Adresse gerichtet, sondern an die Adresse von R3. Unicast-IP-Adressen werden zusammen mit den Befehl/Response-Flags verwendet. Sobald der Austausch abgeschlossen ist, kehren die Router zur Multicast-Adresse zurück, die von Routern gelesen wird, die möglicherweise neu in das Netzwerk eingefügt wurden.

Nachdem R3 ebenfalls für RIPv2 konfiguriert wurde, werden die Routingtabellen über die Request/Response-Pakete aufgefüllt, wie in Tabelle 5-2 zu sehen.

Tabelle 5-2: Vollständig gefüllte Routingtabelle nach RIP

R1	R2	R3
C 192.168.1.0 F0/0	C 192.168.2.0 F0/0	C 192.168.3.0 F0/0
C 192.168.2.0 F0/1	C 192.168.3.0 F0/1	C 192.168.4.0 F0/1
R 192.168.3.0 [120/1] via 192.168.2.254	R 192.168.1.0 [120/1] via 192.168.2.253	R 192.168.1.0 [120/2] via via 192.168.3.253
R 192.168.4.0 [120/2] via 192.168.2.254	R 192.168.4.0 [120/1] via 192.168.3.254	R 192.168.2.0 [120/1] via via 192.168.3.253

Alle Details in der Routingtabelle sind wichtig, doch einige Dinge sind besonders erwähnenswert. Die administrative Distanz (AD) und die Metrik sind in den eckigen Klammern zu finden. RIP hat die administrative Distanz 120, und die Metrik entspricht der Anzahl der Hops. In unserem kleinen Netzwerk ist 2 die größte Metrik. Diese Information kann in den Quell-RIP-Paketen wie in Abbildung 5-2 bis Abbildung 5-4 verfolgt werden.

Ein weiteres wichtiges Detail ist der Forwarding-Router, oder der nächste Hop. In der Routingtabelle ist das die »via«-Adresse. Diese Adresse lernt der Router anhand der Quell-IP-Adresse im RIP-Paket. Wie Sie sehen, besitzen einige der über RIP gelernten Routen die gleiche Forwarding-Adresse. Zum Beispiel sendet R3 Traffic für die Netzwerke 192.168.1.0 und 192.168.2.0 an 192.168.3.253. Das ist so, wie es für diese Topologie sein sollte, doch wie im Abschnitt über statische Routen in Kapitel 1 beschrieben, ist das ein möglicher Kandidat für eine Default-Route. Die tatsächliche Routingtabelle für R1 ist in Abbildung 5-10 zu sehen.

▼ Abbildung 5-10
Tatsächliche Routingtabelle von R1

```
R1#
R1#show ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is not set

R    192.168.4.0/24 [120/2] via 192.168.2.254, 00:00:13, FastEthernet0/1
C    192.168.1.0/24 is directly connected, FastEthernet0/0
C    192.168.2.0/24 is directly connected, FastEthernet0/1
R    192.168.3.0/24 [120/1] via 192.168.2.254, 00:00:13, FastEthernet0/1
R1#
```

Diese Ausgabe wurde vom Befehl `show ip route` generiert. Der Router weist jedem dynamischen Eintrag auch eine Zeit zu. Auf diese Weise lässt sich das Alter erlernter Routen nachhalten.

Timer

Wie viele andere Protokolle besitzt RIP eine Reihe von Timern, die das Ankündigen und Entfernen alter oder falscher Routinginformationen steuern.

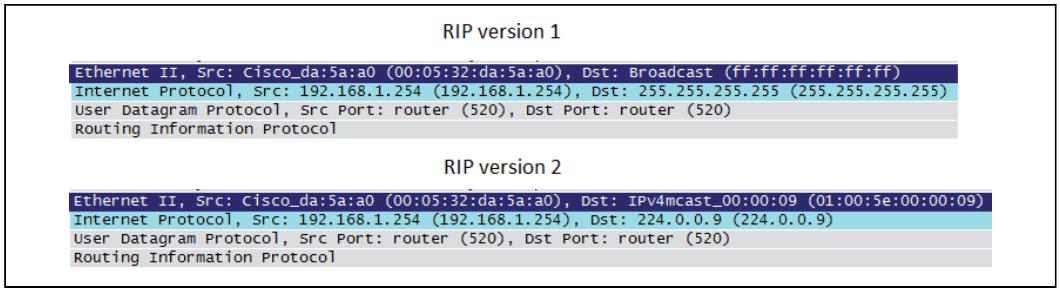
Response- oder Update-Timer
: Während des normalen Betriebs sendet der Routingprozess alle 30 Sekunden unverlangt eine Response-Meldung, um seine Routinginformationen aktuell zu halten.

Route-Timeout oder Ungültig-(invalid-)Timer
: Nach 180 Sekunden wird jeder Router, der nicht über ein Response-Paket aktualisiert wurde, als fehlerhaft betrachtet und aus der Routingtabelle entfernt. Nach Ablauf dieser Zeitspanne werden die benachbarten Router über Updates informiert, dass diese Route fehlerhaft ist, und der Garbage Collection-Timer wird gesetzt. In den Updates wird die Metrik für dieses Ziel auf 16 gesetzt.

Garbage Collection- oder Flush-Timer
: Läuft dieser Timer aus, wird die Route schließlich aus der Routingtabelle entfernt. An dieser Stelle können die Implementierungen etwas schwierig sein. RFC 2453 legt fest, dass diese Zeit mit 120 Sekunden eingestellt werden soll. Cisco verwendet 60 Sekunden von dem Zeitpunkt an, wenn dieser Timer ausläuft, bzw. 240 Sekunden ingesamt für den Routen-Eintrag. Cisco verweist auf einen »Hold Down«-Timer, der den Zeitunterschied erklärt. Die Dokumentation spricht allerdings von 180 Sekunden.

Adressierung

Ein weiteres wichtiges Detail findet sich nicht in der Routingtabelle, sondern in der Adressierung in den Headern des RIP-Pakets. Abbildung 5-11 zeigt sowohl RIPv1- als auch RIPv2-Pakete.



▲ **Abbildung 5-11**
RIP-Adressierung

Beide Pakete besitzen eine Quell-IP-Adresse, die dem sendenden Interface des Routers entspricht. Allerdings verwendet die RIP-Version 1 eine beschränkte Broadcast-Adresse (255.255.255.255) als Ziel, während die Version 2 die reservierte Multicast-Adresse 224.0.0.9 nutzt. Die Schicht-2-Adressierung folgt häufig der Schicht-3-Adressierung, und daher verwendet das RIPv1-Paket eine Broadcast-Adresse für den Ethernet-Frame. Das RIPv2-Paket verwendet eine Multicast-MAC-Adresse im Schicht-2-Frame, die auf der Schicht-3-IP-Multicast-Adresse basiert.

Zwar geht es in diesem Kapitel nicht um die Multicast-Adressierung, aber ein wenig Hintergrundwissen ist doch hilfreich. Tabelle 5-3 zeigt die allgemeine Multicast-Adressierung, wie sie in RFC 3171 beschrieben wird:

Tabelle 5-3: RFC 3171-Multicast-Adressierung

Adresse	Zweck
224.0.0.0 - 224.0.0.255	Local Network Control Block
224.0.1.0 - 224.0.1.255	Internetwork Control Block
224.0.2.0 - 224.0.255.0	AD-HOC Block
224.1.0.0 - 224.1.255.255	ST Multicast Groups
224.2.0.0 - 224.2.255.255	SDP/SAP Block
224.252.0.0 - 224.255.255.255	DIS Transient Block
225.0.0.0 - 231.255.255.255	RESERVED
232.0.0.0 - 232.255.255.255	Source Specific Multicast Block
233.0.0.0 - 233.255.255.255	GLOP Block
234.0.0.0 - 238.255.255.255	RESERVED
239.0.0.0 - 239.255.255.255	Administratively Scoped Block

Innerhalb des »Local Network Control Blocks« gibt es verschiedene Adressen, die uns besonders am Herzen liegen:

224.0.0.1
: Alle Hosts-Multicast

224.0.0.2
: Alle Router-Multicast

224.0.0.5
: OSPF

224.0.0.9
: RIPv2

Diese Adresse wird RIPv2 vom RFC zugewiesen. Da Router die einzigen Geräte sind, die RIPv2 üblicherweise ausführen, verarbeiten andere Geräte diese Pakete normalerweise nicht. Multicasting kann für Netzwerkadministratoren eine echte Herausforderung sein, da die Router Multicast-Pakete nicht weiterleiten, zumindest nicht ohne Hilfe des protokollunabhängigen Multicastings (Protocol Independent Multicast, PIM) und des Interior Group Management Protocol (IGMP). Glücklicherweise werden RIPv2-Pakete nicht einfach weitergeleitet. Sie werden verändert und dann erneut gesendet.

Der letzte Teil der im Paket enthaltenen Adressierung ist eine Schicht-4-UDP-Portnummer. Sowohl RIPv1 als auch RIPv2 verwenden Port 520. Es ist manchmal lustig zu beobachten, wie neue Netzwerkadministratoren Zugriffskontrolllisten oder Firewall-Regeln konfigurieren. Sie sind so damit beschäftigt, unerwünschten UDP/TCP-Traffic zu blockieren, dass RIP manchmal herausgefiltert wird, woraufhin der Administrator sich wundert, warum es so viele ICMP-»Destination Unreachable«-Meldungen gibt.

Fortgeschrittener Betrieb

Den einfachen RIP-Betrieb kann man verstehen, wenn man sich die Pakete ansieht. RIP-Pakete sind auch sehr aufschlussreich in Hinblick darauf, was sie *nicht* enthalten. In diesem Abschnitt wollen wir uns einige zusätzliche Regeln ansehen, die in das Protokoll integriert wurden, um Probleme zu vermeiden.

Split Horizon

Wenn Sie zwei Personen beobachten, die sich einander vorstellen, würden Sie eine Unterhaltung wie die folgende erwarten:

Person 1: »Hallo, mein Name ist Bob.«
Person 2: »Hallo, mein Name ist Sally.«

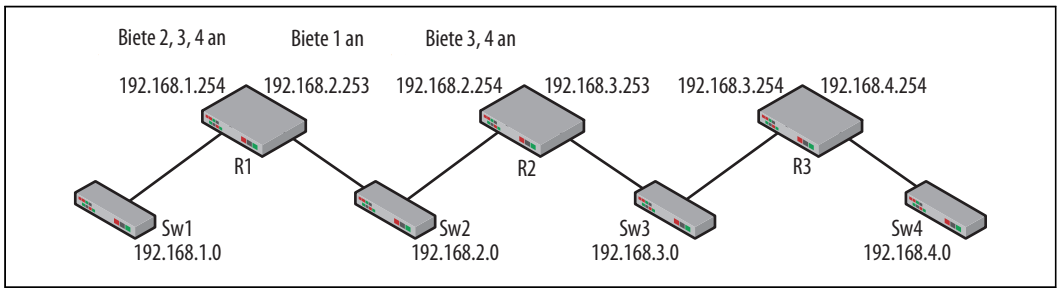
Sie würden NICHT das hier erwarten:

Person 1: »Hallo, mein Name ist Bob.«
Person 2: »Hallo, dein Name ist Bob.«

Bob weiß bereits, dass sein Name Bob ist, und daher ist es albern von Sally, Bob etwas zu sagen, das er gerade gesagt hat. Das Gleiche gilt auch für Router, d.h., Router sollen ihren Nachbarn nichts über Netzwerke erzählen, die diese bereits kennen. Anders ausgedrückt: Sie sollten nichts über die Schnittstelle anbieten, über die Sie es gelernt haben. Es gibt auch keinen Grund, warum man die Verfügbarkeit des Netzwerks eben diesem Netzwerk bekannt geben sollte.

In Abbildung 5-12 ist R1 direkt mit 192.168.1.0 und 192.168.2.0 verbunden. R1 zeigt das 192.168.1.0-Netzwerk dem 192.168.1.0-Netzwerk nicht an. Das Gleiche gilt auch für R2 im 192.168.2.0-Netzwerk.

▼ **Abbildung 5-12**
Split Horizon-Advertising



Wir können das Verhalten zwischen R1 und R2 genauer skizzieren. R1 bietet rechts das Netzwerk 192.168.1.0 an und links die Netzwerke 192.168.2.0, 192.168.3.0 und 192.168.4.0. R2 empfängt Informationen zum Netzwerk 192.168.1.0 von R1 und ist direkt mit dem Netzwerk 192.168.2.0 verbunden. R1 erhält also nur die Netzwerke 192.168.3.0 und 192.168.4.0 zurück. Der Split Horizon-Betrieb ist in den Paketen zu sehen. Abbildung 5-13 zeigt die Pakete von R2 und R3, wie sie im 192.168.2.0-Netzwerk erscheinen.

▼ **Abbildung 5-13**
Split Horizon-Paketvergleich

```
Ethernet II, Src: Cisco_da:5a:a1 (00:05:32:da:5a:a1), Dst: IPv4mcast_00:00:09 (01:00:5e:00:00:09)
Internet Protocol, Src: 192.168.2.253 (192.168.2.253), Dst: 224.0.0.9 (224.0.0.9)
User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
Routing Information Protocol
  Command: Response (2)
  Version: RIPv2 (2)
  Routing Domain: 0
⊞ IP Address: 192.168.1.0, Metric: 1

Ethernet II, Src: Cisco_28:02:80 (00:05:5e:28:02:80), Dst: IPv4mcast_00:00:09 (01:00:5e:00:00:09)
Internet Protocol, Src: 192.168.2.254 (192.168.2.254), Dst: 224.0.0.9 (224.0.0.9)
User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
Routing Information Protocol
  Command: Response (2)
  Version: RIPv2 (2)
  Routing Domain: 0
⊞ IP Address: 192.168.3.0, Metric: 1
⊞ IP Address: 192.168.4.0, Metric: 2
```

Die IP-Adressen in diesen Paketen zeigen, dass sie von R2 und R3 stammen. Wir sehen, dass die Router die Split Horizon-Regeln anwenden und auf diese Weise die Größe der Pakete minimieren. Doch der eigentliche Vorteil von Split Horizon ist die schnellere Konvergenz, da die Pfade zu den Zielen klar sind.

Was passiert, wenn Split Horizon nicht genutzt wird? Wie sich zeigt, verwenden einige WAN-Verbindungen Split Horizon nicht, doch das ist eher selten. Split Horizon zu deaktivieren ist üblicherweise keine gute Sache. Nehmen wir an, dass die Router in der obigen Topologie alle Netzwerke auf allen Interfaces anbieten, wie in Abbildung 5-14 zu sehen ist. Um das Ausmaß des Problems zu verdeutlichen, wurde ein weiterer Router in die Netzwerke eingefügt.

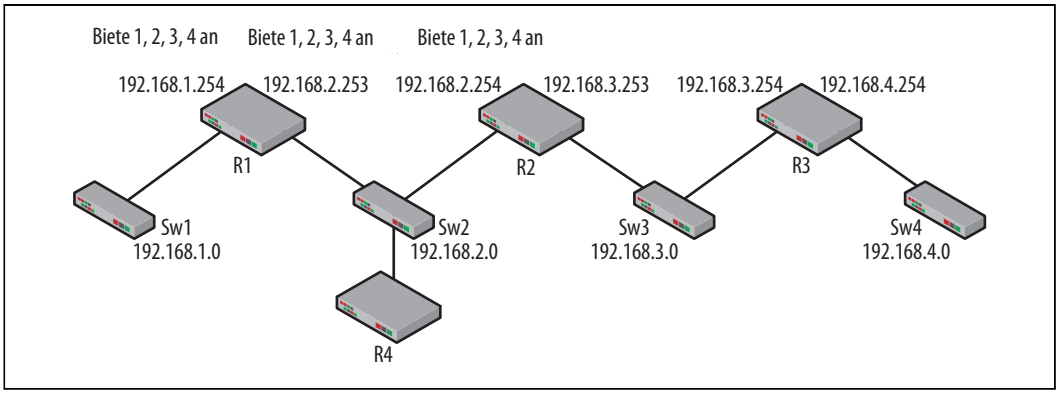


Abbildung 5-14 ▲ Advertisements ohne Split Horizon

Nehmen wir nun an, dass R1 ausfällt. R1 war der einzige Weg ins 192.168.1.0-Netzwerk. Allerdings zeigt R2 die Verfügbarkeit des 192.168.1.0-Netzwerks an. Tatsächlich zeigt R4 das Netzwerk auch an, wenn R4 Split Horizon nicht nutzt. Denken Sie daran, dass das Netzwerk 192.168.1.0 nicht länger verfügbar ist. Alle Router in der Topologie glauben aber, dass dieses Netzwerk noch verfügbar ist, und halten es in ihren Routingtabellen vor. Ein anderes mögliches Szenario ist, dass nicht R1 verloren geht, sondern nur die Schnittstelle 192.168.1.254. R1 würde 192.168.1.0 nicht mehr anbieten, doch nach den Informationen von R2 würde er glauben, dass das Netzwerk von der anderen Seite der Topologie zu erreichen ist. Split Horizon ist standardmäßig aktiviert, um solche Konvergenzprobleme zu vermeiden.

Poisoning

Einen weiteren Schutz bietet das sogenannte »Vergiften« (Poisoning) von Routen. Falls sich die Router-Konfiguration ändert oder ein Gerät ausfällt, kann ein Router die Route vergiften, so dass andere Router wissen, dass diese(s) Netzwerk(e) nicht mehr verfügbar ist/sind. Eine Route wird vergiftet, indem der Router eine Metrik einfügt, die »unendlich« ist. Für RIP ist das der Wert 16.

Was würde passieren, wenn bei der gleichen Topologie 192.168.3.253 die Verbindung zum Netzwerk 192.168.3.0 verliert? Solange R2 immer noch über 192.168.2.254 verbunden ist, kann er das 192.168.3.0-Netzwerk vergiften. Router, die ein in dieser Form vergiftetes Paket empfangen, wissen sofort, dass dieser Weg nicht mehr zur Verfügung steht, und entfernen ihn schneller aus der Routingtabelle. Ein auf diese Weise vergiftetes Paket ist in Abbildung 5-15 zu sehen.

```
Ethernet II, Src: Cisco_28:02:80 (00:05:5e:28:02:80), Dst: IPv4mcast_00:00:09 (01:00:5e:00:00:09)
Internet Protocol, Src: 192.168.2.254 (192.168.2.254), Dst: 224.0.0.9 (224.0.0.9)
User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
Routing Information Protocol
  Command: Response (2)
  Version: RIPv2 (2)
  Routing Domain: 0
⊟ IP Address: 192.168.3.0, Metric: 16
    Address Family: IP (2)
    Route Tag: 0
    IP Address: 192.168.3.0 (192.168.3.0)
    Netmask: 255.255.255.0 (255.255.255.0)
    Next Hop: 0.0.0.0 (0.0.0.0)
    Metric: 16
```

▲ **Abbildung 5-15**
»Vergiftetes« Paket

Würde R2 komplett ausfallen, müssten die anderen Router in der Topologie auf ihre Timer bauen, um das Problem zu lösen. Das Poisoning von Routen ist standardmäßig aktiv.

Poison Reverse

Poison Reverse baut auf der Idee des Poisonings auf, wird aber während des normalen (stabilen) Betriebs genutzt, um sicherzustellen, dass kein Versuch unternommen wird, ein Netzwerk über eine ungeeignete oder unerwünschte Route zu erreichen. In obiger Topologie, in der R1 die Verfügbarkeit des Netzwerks 192.168.1.0 anzeigt, gibt R2 die Unerreichbarkeit des gleichen Netzwerks an R1 zurück. Passiert nun irgendetwas mit R1, verkünden die anderen Router explizit, dass sie keinen Pfad zu den verlorenen Netzwerken besitzen (siehe Abbildung 5-16).

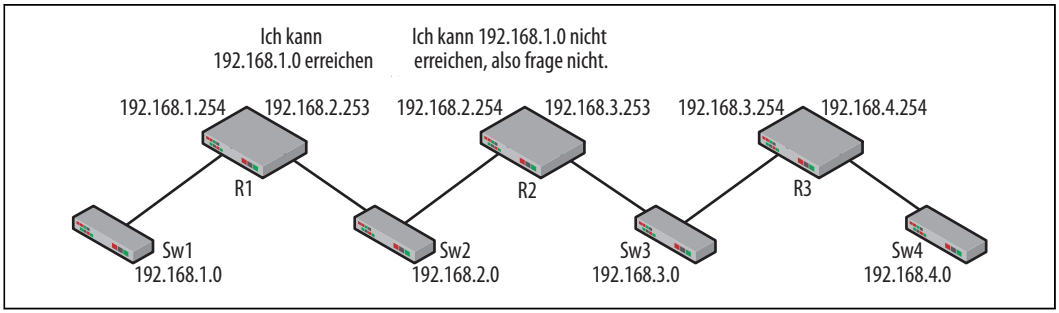


Abbildung 5-16 ▲
Poison Reverse-Benachrichtigungen

Poison Reverse ist standardmäßig nicht aktiv und muss daher auf dem Router aktiviert werden. Allerdings verwenden einige Routingimplementierungen Poison Reverse während der »Entdecke deine Nachbarn«-Phase. Abbildung 5-17 zeigt die Request- und Response-Pakete, die zwischen R2 und R3 über das 192.168.2.0-Netzwerk laufen. Obwohl es nicht Teil des normalen RIP-Traffics ist, können wir sehen, dass R1 (192.168.2.253) unmittelbar nach dem Erlernen der 192.168.3.0- und 192.168.4.0-Netzwerke von R2, Poison Reverse nutzt, um R2 darüber zu informieren, dass es keine anderen Pfade zu diesen Zielen kennt. Nach diesem Datenaustausch kehren die RIP-Pakete zur Normalität zurück.

Abbildung 5-17 ▼
Paketaustausch bei Poison Reverse

No.	Source	Destination	Protocol	Time	Info
6	192.168.2.253	255.255.255.255	RIPv1	2.770111	Request
11	192.168.2.253	255.255.255.255	RIPv1	7.452966	Response
12	192.168.2.254	255.255.255.255	RIPv1	8.454928	Request
25	192.168.2.254	255.255.255.255	RIPv1	11.567494	Request
26	192.168.2.253	192.168.2.254	RIPv1	11.569164	Response
32	192.168.2.254	255.255.255.255	RIPv1	19.131660	Response
39	192.168.2.253	255.255.255.255	RIPv1	36.657248	Response

```
⊞ Frame 26: 106 bytes on wire (848 bits), 106 bytes captured (848 bits)
⊞ Ethernet II, Src: Cisco_da:5a:a1 (00:05:32:da:5a:a1), Dst: Cisco_28:02:80 (00:05:5e:28:02:80)
⊞ Internet Protocol, Src: 192.168.2.253 (192.168.2.253), Dst: 192.168.2.254 (192.168.2.254)
⊞ User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
⊟ Routing Information Protocol
    Command: Response (2)
    Version: RIPv1 (1)
  ⊟ IP Address: 192.168.1.0, Metric: 1
      Address Family: IP (2)
      IP Address: 192.168.1.0 (192.168.1.0)
      Metric: 1
  ⊟ IP Address: 192.168.3.0, Metric: 16
      Address Family: IP (2)
      IP Address: 192.168.3.0 (192.168.3.0)
      Metric: 16
  ⊟ IP Address: 192.168.4.0, Metric: 16
      Address Family: IP (2)
      IP Address: 192.168.4.0 (192.168.4.0)
      Metric: 16
```

Getriggerte Updates

Sobald sich Informationen bezüglich der Routingtabelle ändern, sendet der Router ein RIP-Paket mit eben diesen neuen Informationen, ohne darauf zu warten dass der Update-Timer abläuft. Dieses schnelle RIP-Paket wird als getriggertes Update bezeichnet. Die Idee ist, dass sich Informationen zu Fehlern und Änderungen deutlich schneller durch das Netzwerk verbreiten, wenn sie nicht darauf warten, dass die Standard-Update-Timer ablaufen. Router, die solche getriggerten Updates empfangen, können ihrerseits selbst getriggerte Updates senden. Auf diesem Weg erreicht eine Welle neuer Informationen alle Ecken des Netzwerks, was wiederum die Konvergenzgeschwindigkeit erhöht.

Eine Reihe von Beispielen für getriggerte Updates kann man sehen, wenn die Timer ablaufen. Nehmen wir an, die Verbindung zwischen R3 und S3 geht verloren, d.h., R2 empfängt keine Updates mehr von R3 zum 192.168.4.0-Netzwerk. Nach 180 Sekunden wird die Route in der Routingtabelle als möglicherweise nicht mehr erreichbar gekennzeichnet (siehe Abbildung 5-18), und ein getriggertes Update wird gesendet, in dem das 192.168.4.0-Netzwerk die Metrik 16 hat. Diese getriggerten Updates werden nahezu direkt durch das gesamte Netzwerk propagiert. Nach weiteren 60 Sekunden wird die Route aus der Routingtabelle entfernt. Ein anderes Beispiel ist das Herunterfahren der 192.168.4.254-Schnittstelle. In diesem Fall würde sofort ein Update gesendet werden.

▲ **Abbildung 5-18**
192.168.4.0 ist möglicherweise unten.

Getriggerte Updates werden auch bei Verbesserungen gesendet. Ist das 192.168.4.254-Interface wieder oben, werden umgehend getriggerte Updates gesendet und durch das gesamte Netzwerk propagiert. Die Routingtabellen benachbarter Router werden auch sofort aktualisiert.

Count to Infinity

»Count to Infinity« (Zählen bis Unendlich) ist ein weiteres Werkzeug, um das Netzwerk von seinen Fesseln zu befreien, wenn es keine Updates oder »vergifteten« Routen gibt. Es ist das letzte Mittel

bei Konnektivitätsverlust oder Gerätefehlern. Beispielweise ging in Abbildung 5-16 die Verbindung von R3 zu Switch 3 verloren. R2 würde das Problem nicht erkennen, da der Link-Impuls für 192.168.3.253 immer noch vorhanden wäre.

Abbildung 5-19 zeigt eine etwas komplexere Topologie. Es wurde eine Schleife aufgebaut, die dazu führt, dass die Routinginformationen in zwei Richtungen fließen. R4 zeigt die Verfügbarkeit des 192.168.4.0-Netzwerks an und sagt, dass es einen Hop entfernt ist.

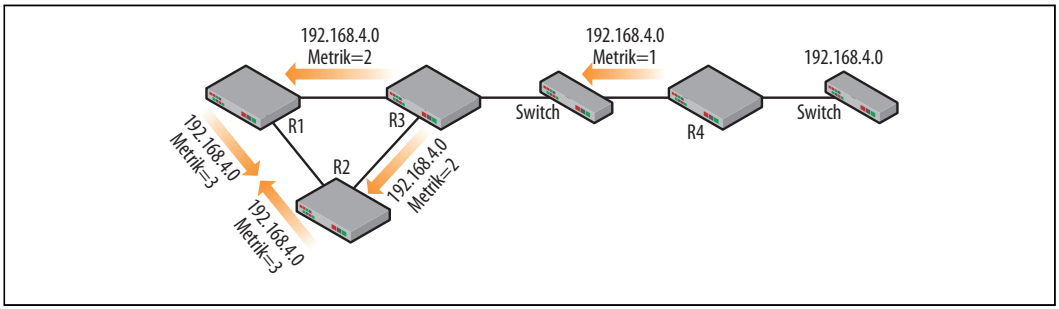


Abbildung 5-19 ▲
Count to Infinity-Problem

R3 bietet dann das gleiche Netzwerk an, nachdem es den Hop-Zähler um eins erhöht hat. Da R3 mit den Downstream-Routern R1 und R2 verbunden ist, treffen dort die gleichen RIP-Informationen ein, wenn auch über verschiedene Interfaces. R1 und R2 bieten sich das Netzwerk gegenseitig an, nachdem sie den Hop-Zähler erhöht haben. Beim Empfang dieser RIP-Pakete entfernen R1 und R2 die Information wieder, da diese schlechter sind als die bereits bekannten Routen.

Doch was passiert beim katastrophalen Verlust von R4? Selbst wenn wir davon ausgehen, dass Split Horizon, Poisoning und getriggerte Updates sauber arbeiten, helfen sie uns in diesem Fall nicht weiter. R3 weiß nicht, dass R4 nicht erreichbar ist, und kann sich nur auf die bereits bekannten Informationen und auf die RIP-Timer verlassen. Irgendwann entfernt R3 die Route und bietet sie auch nicht mehr an. Sobald das passiert, müssen sich die Downstream-Router R1 und R2 um Split Horizon keine Gedanken mehr machen und *beginnen damit, das 192.168.4.0-Netzwerk anzubieten*. Doch die Metrik hat sich erhöht. R3 bietet die Route der anderen Seite des Netzwerks an, nachdem es den Hop-Zähler um eins erhöht hat. Ursprünglich haben R1 und R2 das Netzwerk 192.168.4.0 von R3 gelernt. Aus ihrer Sicht kann sich die Distanz zum Ziel (Metrik) erhöht haben, die Quell-IP-Adresse (Vektor) aber nicht. Also erhöhen sie den Hop-Zähler und senden die RIP-Pakete erneut. Das setzt sich fort, bis das RIP-Paket den Hop-Wert 16 enthält und der Pfad als unbrauchbar betrachtet wird.

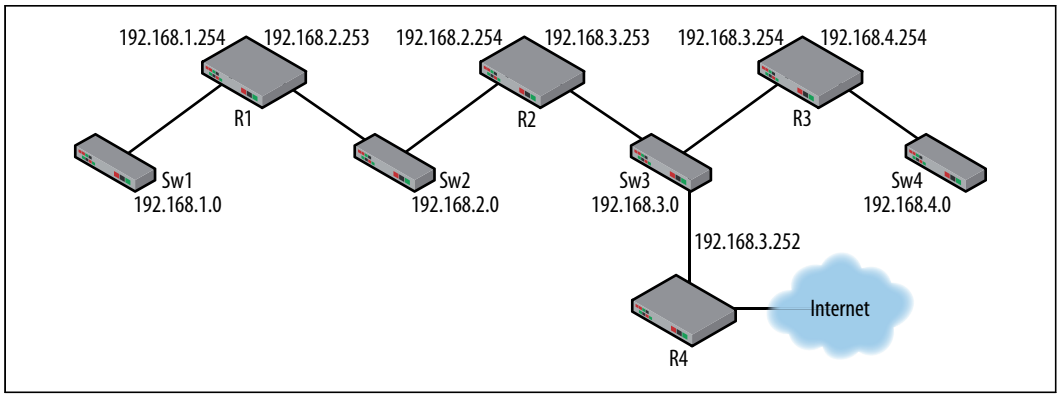
Die Hoffnung ist, dass das Poisoning Routen löscht, getriggerte Updates das Problem lösen und Netzwerkadministratoren sich nie auf diesen zeitaufwendigen Prozess verlassen müssen. Doch RFC 2453 warnt:

> Würde ein System stillhalten, während die Kaskade getriggerter Updates rollt, käme es bewiesenermaßen nie zu einem Count to Infinity. Fehlerhafte Routen würden immer direkt entfernt werden, und es könnte nie zu Routingschleifen kommen. Unglücklicherweise stellen sich die Dinge nicht so schön dar. Während getriggerte Updates gesendet werden, können gleichzeitig auch reguläre Updates auftreten. Router, die die getriggerten Updates noch nicht erhalten haben, senden weiterhin Informationen, die auf Routen basieren, die es nicht mehr gibt. Es ist möglich, dass nach einem getriggerten Update ein Router ein normales Update von einem dieser Router empfängt, der noch keine aktuellen Daten hat. Damit könnte die fehlerhafte Route wiederhergestellt werden.

Wie komme ich aus einem Netzwerk raus?

Bis zu diesem Punkt haben wir RIP verwendet, um Ziele innerhalb der Gruppe RIP-basierter Netzwerke bzw. dem, was das RFC als autonomes System bezeichnet, zu erreichen. Doch irgendwo anders kann der Traffic nicht hin. Wie erfolgt also der Übergang vom internen Gateway-Protokoll zum Rest der Welt? In Kapitel 1 wurde das Routing im Allgemeinen behandelt, und es gab einen Abschnitt über die Default-Route. Da die Topologie in diesem Kapitel genau die gleiche war, gelten auch die gleichen Regeln. Ein Kandidat für die Default-Route kommt üblicherweise mehrmals in den Routingtabellen anderer Router vor. Modifiziert man die Topologie ein wenig, führt ein klarer Pfad aus dieser Gruppe von Netzwerken heraus (siehe Abbildung 5-20).

▼ **Abbildung 5-20**
Topologie für RIP-Default-Route



Trotz des Hinzufügens von R4 bleibt die Topologie einfach. Einerseits kann der Netzwerkadministrator einfach Default-Routen auf allen Routern installieren. Allerdings wäre das Netzwerk dann vor Änderungen in der Topologie oder unterbrochenen Verbindungen nicht geschützt.

Eine andere Strategie, die Sie mit RIP verwenden können, ist die Redistribution. Als Weg nach draußen kann RIP eine Default-Route installieren, die ins Internet zeigt. Betreibt man RIP auf der 192.168.3.0-Seite, kann diese Default-Route an die Downstream-Server (R1, R2, R3) mithilfe des Befehls `redistribute static` weiterverteilt werden. Die Grundkonfiguration von R4 sieht wie folgt aus:

```
router rip
    version 2
    redistribute static
    network 192.168.3.0
ip route 0.0.0.0 0.0.0.0 10.101.100.254
```

Sobald der `redistribute`-Befehl abgesetzt wurde, laufen RIP-Pakete (samt Default-Route) flussabwärts. R1, R2 und R3 aktualisieren ihre Routingtabellen und binden diese neuen Informationen ein. Ein solches Paket ist in Abbildung 5-21 zu sehen.

```
⊞ Ethernet II, Src: Cisco_28:1c:a0 (00:05:5e:28:1c:a0), Dst: IPv4mcast_00:00:09 (01:00:5e:00:00:09)
⊞ Internet Protocol, Src: 192.168.3.252 (192.168.3.252), Dst: 224.0.0.9 (224.0.0.9)
⊞ User Datagram Protocol, Src Port: router (520), Dst Port: router (520)
⊟ Routing Information Protocol
    Command: Response (2)
    Version: RIPv2 (2)
    Routing Domain: 0
  ⊟ IP Address: 0.0.0.0, Metric: 1
      Address Family: IP (2)
      Route Tag: 0
      IP Address: 0.0.0.0 (0.0.0.0)
      Netmask: 0.0.0.0 (0.0.0.0)
      Next Hop: 0.0.0.0 (0.0.0.0)
      Metric: 1
```

Abbildung 5-21 ▲
RIP-Paket mit Default-Route

Abbildung 5-22 verdeutlicht die Änderungen in den Routingtabellen. Beachten Sie, dass R2 und R3 mit dem gleichen Netzwerk verbunden sind wie R4 und dass sie R4 als Default-Route mit einem Hop-Wert von 1 verwenden. Doch das RIP-Paket wurde durch R2 aktualisiert, und R1 verwendet nun R2 als Standard-Gateway (mit einem höheren Hop-Wert).

```
Gateway of last resort is 192.168.2.254 to network 0.0.0.0

R    192.168.4.0/24 [120/2] via 192.168.2.254, 00:00:06, FastEthernet0/1
C    192.168.1.0/24 is directly connected, FastEthernet0/0
C    192.168.2.0/24 is directly connected, FastEthernet0/1
R    192.168.3.0/24 [120/1] via 192.168.2.254, 00:00:06, FastEthernet0/1
R*   0.0.0.0/0 [120/2] via 192.168.2.254, 00:00:06, FastEthernet0/1
R1#
```

```
Gateway of last resort is 192.168.3.252 to network 0.0.0.0

R    192.168.4.0/24 [120/1] via 192.168.3.254, 00:00:09, FastEthernet0/1
R    192.168.1.0/24 [120/1] via 192.168.2.253, 00:00:25, FastEthernet0/0
C    192.168.2.0/24 is directly connected, FastEthernet0/0
C    192.168.3.0/24 is directly connected, FastEthernet0/1
R*   0.0.0.0/0 [120/1] via 192.168.3.252, 00:00:05, FastEthernet0/1
R2#
```

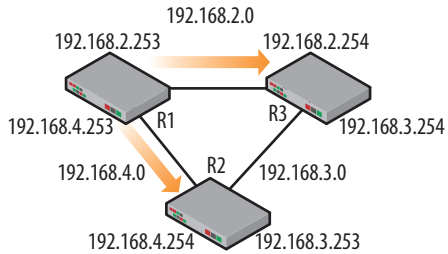
```
Gateway of last resort is 192.168.3.252 to network 0.0.0.0

C    192.168.4.0/24 is directly connected, FastEthernet0/1
R    192.168.1.0/24 [120/2] via 192.168.3.253, 00:00:24, FastEthernet0/0
R    192.168.2.0/24 [120/1] via 192.168.3.253, 00:00:24, FastEthernet0/0
C    192.168.3.0/24 is directly connected, FastEthernet0/0
R*   0.0.0.0/0 [120/1] via 192.168.3.252, 00:00:00, FastEthernet0/0
R3#
```

▲ **Abbildung 5-22**
Routingtabellen mit installierten Default-Routen

RIP und Schleifen

Routingschleifen können durch physische Verbindungen oder eine falsche Konfiguration entstehen. Schleifen enthaltende Architekturen können die Datenübertragung massiv behindern. Die meisten Protokolle, inklusive RIP, wenden Techniken (wie sie vorhin beschrieben wurden) an, um die Auswirkungen von Schleifen auf IP-Pakete zu beschränken. Doch was passiert, wenn eine Schleife in die Topologie eingefügt wird? Abbildung 5-23 zeigt R1, R2 und R3, die in einer Schleife miteinander verbunden sind. Das Netzwerk 192.168.1.0 wurde entfernt, und R1 wurde eine Adresse im 192.168.4.0-Netzwerk zugewiesen. In einer solchen Topologie fließen die RIP-Pakete genau so wie bei den bisher diskutierten Topologien auch.

Abbildung 5-23 ▶
Topologie mit Schleife



Untersucht man die Topologie aus der Perspektive von R1, ist er direkt mit den Netzwerken 192.168.2.0 und 192.168.4.0 verbunden. Er ist außerdem einen Hop vom 192.168.3.0-Netzwerk entfernt. Doch dieses Netzwerk kann aus zwei verschiedenen Richtungen erreicht werden. Der Vorteil besteht darin, dass ein Pfad automatisch übernehmen kann, wenn der andere Weg verloren geht. Die Routingtabelle von R1 ist in Abbildung 5-24 zu sehen.

```
Gateway of last resort is not set

C    192.168.4.0/24 is directly connected, FastEthernet0/0
C    192.168.2.0/24 is directly connected, FastEthernet0/1
R    192.168.3.0/24 [120/1] via 192.168.2.254, 00:00:06, FastEthernet0/1
                    [120/1] via 192.168.4.254, 00:00:15, FastEthernet0/0
R1#
```

Abbildung 5-24 ▲
R1-Routingtabelle

Die auf den beiden direkt verbundenen Netzwerken abgefangenen Pakete zeigen, dass R1 ein Load-Balancing durchführt, wenn Traffic an das 192.168.3.0-Netzwerk gesendet wird. Das heißt, eine Hälfte des Traffics wird über 192.168.2.254 geleitet und die andere über 192.168.4.254.

Sicherheit

Die Netzwerksicherheit umfasst viele Aspekte, inklusive der Sicherheit von Netzwerk-Geräten und den im Netzwerk laufenden Protokollen. Routingprotokolle sind bekanntermaßen leicht zu stören. Wie die getriggerten Updates gezeigt haben, fragt ein Router nicht lange, wenn er neue oder bessere Informationen erhält, sondern assimiliert sie zügig in seine Routingtabelle. Traffic kann daher umgeleitet werden, gute Daten können verdrängt werden und Traffic kann über nicht existierende Pfade geschickt werden. Stellen Sie sich vor, dass ein Eindringling, der sich Zugang zum Netzwerk verschafft

hat, nicht nur den Datenverkehr abfangen, sondern auch neuen Traffic einspeisen kann. Router, die Informationen von einem Angreifer erhalten, können diese nicht von den authentischen Informationen benachbarter Router unterscheiden.

Das Problem kann man auf unterschiedliche Arten angehen. Das Management von Routern kann auf bestimmte Segmente oder Interfaces beschränkt werden. Zusätzlich kann der für die Router gedachte Traffic gefiltert werden. Router achten dann nicht mehr auf Routing-Updates aus einer bestimmten Richtung oder reagieren nicht mehr auf ICMP-Nachrichten und andere Requests. Ein weiteres wertvolles Tool im Werkzeugkasten des Administrators ist das Loopback-Interface. Loopbacks sind Software-Interfaces, die nicht an eine bestimmte physische Schnittstelle gebunden sind. Das bedeutet, dass das Loopback immer verfügbar ist, auch wenn die physischen Ports heruntergefahren sind. Loopbacks können auch IP-Adressen zugewiesen werden, die sich von denen den Netzwerks unterscheiden, so dass ein Angreifer auf das Management-Interface des Gerätes nicht zugreifen kann. Zu guter Letzt können Routingprotokolle auf den Loopback-Schnittstellen laufen. Diese Techniken können zusammen das Management-Netzwerk effektiv vom Datennetzwerk trennen.

RIPv2 besitzt eine weitere Fähigkeit, die das Leben eines Angreifers etwas erschwert: die Authentifizierung von RIP-Nachrichten. Setzt man (wie weiter oben erwähnt wurde) die RIP-AFI auf FFFF, dann wird die Nachricht zur Authentifizierung der restlichen Informationen der RIP-Response verwendet. Die Authentifizierungsdaten (Credentials) werden auf jedem Router innerhalb der AS-Topologie konfiguriert. RFC 2453 legt fest, dass die Authentifizierung über ein einfaches Passwort im Klartext erfolgt, während RFC 2082 eine MD5-basierte Authentifizierung empfiehlt. Beide wurden durch RFC 4822 aktualisiert, das zusätzliche Schlüssel-Algorithmen unterstützt. Einer der größten Unterschiede im Update besteht darin, dass das RIPv2-Paket so modifiziert wird, dass es die Authentifizierungsinformationen an das Ende des Pakets anhängt und nicht einfach in die Felder für das Netzwerkziel einträgt.

RIP und IPv6

Es gibt ein Einsatzmodell für IPv6 RIP. IPv6 RIP ist auch als RIPng oder als »RIP next generation « bekannt. RFC 2080 zeigt auf, dass die Struktur und der Betrieb sich nicht wesentlich von der IPv4-Konfiguration unterscheiden. Einige Änderungen sind hier aufgeführt. Abbildung 5-25 zeigt eine Topologie, die der bisher verwen-

deten recht ähnlich ist. Die Router-Schnittstellen wurden für IPv6-Adressen umkonfiguriert. Eine Erläuterung des statischen Routings der gleichen IPv6-Topologie finden Sie in Kapitel 1.

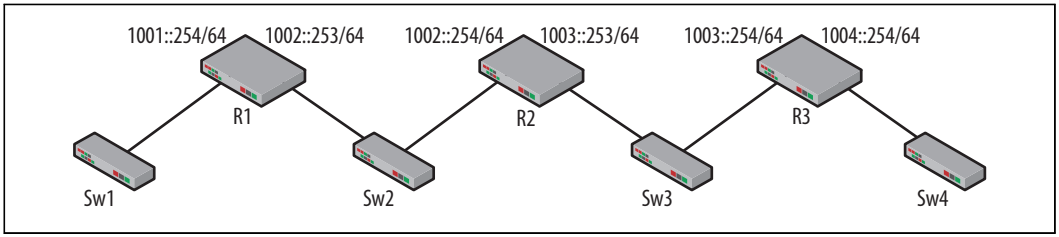


Abbildung 5-25 ▲
IPv6-Topologie

Die grundlegende IPv6-Konfiguration des Routers ist einfach, zeigt aber einen wesentlichen Unterschied: Die RIP-Befehle sind an die Schnittstelle gekoppelt. Das Wort »book« ist einfach ein Schlüsselwort für diese Instanz des RIP-Prozesses.

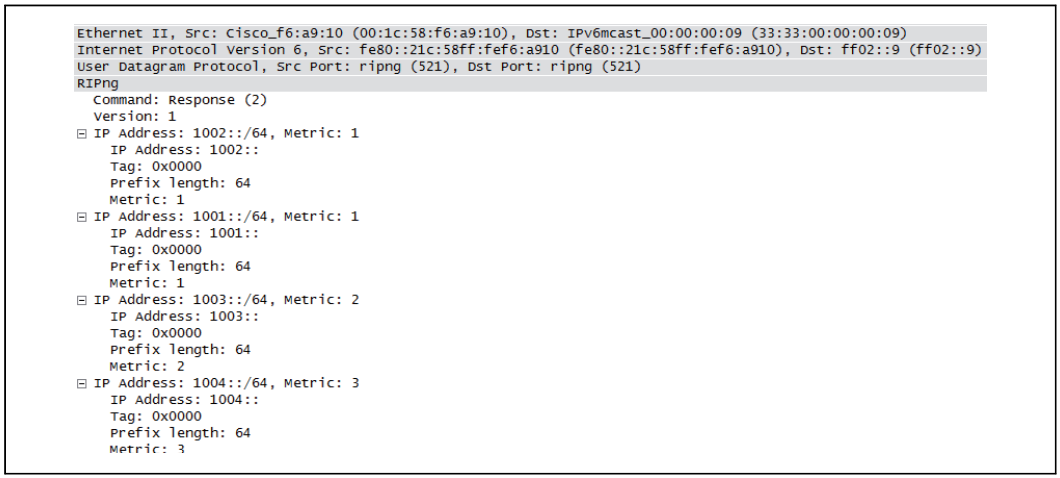
```
ipv6 unicast-routing
interface FastEthernet0/0
    ipv6 address 1001::254/64
    ipv6 rip book enable
!
interface FastEthernet0/1
    ipv6 address 1002::253/64
    ipv6 rip book enable
ipv6 router rip book
```

Abbildung 5-26 zeigt die Routingtabelle für R1. IPv6 fügt die Link-Routen hinzu, und die Netzwerke 1001 und 1002 sind direkt verbunden. Die Netzwerke 1003 und 1004 wurden über RIP erlernt. Beachten Sie, dass die administrative Distanz und die Hop-Zähler auf die gleiche Weise genutzt werden.

Abbildung 5-26 ▶
IPv6-RIP-Routingtabelle

```
C   1001::/64 [0/0]
     via FastEthernet0/0, directly connected
L   1001::254/128 [0/0]
     via FastEthernet0/0, receive
C   1002::/64 [0/0]
     via FastEthernet0/1, directly connected
L   1002::253/128 [0/0]
     via FastEthernet0/1, receive
R   1003::/64 [120/2]
     via FE80::219:2FFF:FE8E:DB48, FastEthernet0/1
R   1004::/64 [120/3]
     via FE80::219:2FFF:FE8E:DB48, FastEthernet0/1
L   FF00::/8 [0/0]
     via Null0, receive
```

Im Einsatz ist IPv6 RIP nahezu identisch, auch wenn Pakete etwas modifiziert wurden, um das andere Schicht-3-Protokoll unterbringen zu können. Darüber hinaus gibt es eine Änderung im Verhalten von Split Horizon. Obwohl IPv6-RIP ihm gehorcht, bietet es das lokale Netzwerk an. Das in Abbildung 5-27 abgebildete Paket wurde im 1001::/64-Netzwerk festgehalten. IPv6 hat eine andere Sicht des Netzwerks und nutzt die Adressierung lokaler Links anstelle global geltender IP-Adressen.



▲ **Abbildung 5-27**
IPv6-RIP-Paket im 1001-Netzwerk

Dieses Paket bietet alle vier bekannten Netzwerke an und nicht nur diejenigen, die über die andere Seite des Routers gelernt wurden. Das Ziel ist eine reservierte IPv6-Multicast-Adresse (FF02::9), und die Portnummer ist 521. Wie man sieht, ist die Struktur sehr ähnlich, und auch wenn sie Version 1 genannt wird, enthält sie Informationen zu Masken und Präfixlängen.

Lektüre

RFC 1058: »Routing Information Protocol«
RFC 1112: »Host Extensions for IP Multicasting«
RFC 1256: »ICMP Router Discovery Messages«
RFC 1812: »Requirements for IP Version 4 Routers«
RFC 1923: »RIPv1 Applicability Statement for Historic Status«
RFC 2080: »RIPng for IPv6«
RFC 2453: »RIP Version 2 (obsoletes 1723, 1388)«
RFC 3171: »IANA Guidelines for IPv4 Multicast Address Assignments«
RFC 4822: »RIPv2 Cryptographic Authentication (obsoletes 2082 RIP-2 MD5 Authentication)«

Zusammenfassung

RIP und Distanzverktor-Routing werden seit den frühesten Tagen der Internet-Kommunikation genutzt. Aufgrund seiner hohen Konvergenzzeit hat RIP eine administrative Distanz von 120. Das macht Routing-Updates von RIP weniger »wünschenswert« als die von anderen Protokollen. Dass es die Hop-Anzahl als Metrik verwendet und da es für eine kleine Gruppe von Netzwerken entworfen wurde, hat RIP eine maximale Netzwerkgröße von 15. RIP hat trotz seiner Mängel überlebt, größtenteils aufgrund einer Reihe von Techniken wie Split Horizon, Poisoning, Poison Reverse, Count to Infinity und getriggerten Updates. Die Unterstützung der Authentifizierung erhöht die Sicherheit des alternden Protokolls und hält RIP weiter am Leben.

Fragen

1. Der Hauptunterschied zwischen RIPv1 und RIPv2 ist die Unterstützung von Subnetzen.
   a. WAHR
   b. FALSCH
2. Welche Metrik wird bei RIP verwendet?
   a. Kosten
   b. Hop-Anzahl
   c. Auslastung
3. Welche administrative Distanz hat RIP?
   a. 90
   b. 100
   c. 110
   d. 120
4. Sowohl RIPv1 als auch RIPv2 verwenden Multicast-Zieladressen.
   a. WAHR
   b. FALSCH

   Weisen Sie den Timern die richtigen Werte zu:
5. Update A. 180
6. Route Timeout B. 120
7. Garbage Collection (basierend auf RFC C. 30)

8. Split Horizon hält Router dazu an, die gesamte Routingtabelle in alle Richtungen weiterzuleiten.
   a. WAHR
   b. FALSCH
9. Bei einem »vergifteten« Paket wird die Metrik auf 16 gesetzt.
   a. WAHR
   b. FALSCH
10. RIP kann nicht in Topologien verwendet werden, die Schleifen enthalten.
   a. WAHR
   b. FALSCH

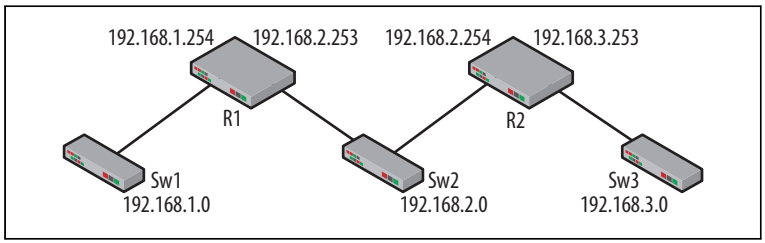
Antworten

1. WAHR
2. B. Hop count
3. D. 120
4. FALSCH
5. C. 30
6. A. 180
7. B. 120
8. FALSCH
9. WAHR
10. FALSCH

Laborübungen

Übung 1: Aufbau der Topologie aus Abbildung 5-28

Material: 2 Router, 2 Computer, optional Switches (oder VLANs) für jedes Netzwerk



◀ **Abbildung 5-28**
Topologie für Übung 1

1. Bauen Sie die Topologie auf, und konfigurieren Sie die IP-Adressen der Router-Interfaces.
2. Verbinden Sie jeweils einen Computer mit den Netzwerken 192.168.1.0 und 192.168.2.0.
3. Konfigurieren Sie manuell die IP-Adressen und Gateways für die Computer.
4. Spielt es für den Computer im 192.168.2.0-Netzwerk eine Rolle, welches Standard-Gateway verwendet wird? Warum? Wie sieht es aus, nachdem RIP läuft?
5. Untersuchen Sie die Routingtabellen auf den Routern. Was enthalten sie? Praktischer Cisco-Befehl: `show ip route`

Übung 2: RIP auf den Routern aktivieren

Material: Topologie aus Übung 1, Wireshark

1. Konfigurieren Sie alle Router für den Einsatz von RIP. Verwenden Sie RIP Version 2.
2. Praktische Cisco-Befehle: `router rip`, `network _____`, `version`
3. Halten Sie den Traffic auf beiden Computern fest, und beobachten Sie, wie die RIP-Pakete anfangen zu fließen. Sobald die Konfiguration abgeschlossen ist, untersuchen Sie erneut die Routingtabellen der Router.
4. Was hat sich in den Routingtabellen der Router geändert? Welche Werte stehen zwischen den eckigen Klammern? Warum?
5. Steht die Tatsache, dass RIP im Netzwerk läuft, in irgendeinem Zusammenhang mit den Host-Routingtabellen?

Übung 3: Split Horizon

Material: Topologie aus Übung 1, Wireshark

1. Was ist Split Horizon? Split Horizon mit Poison Reverse?
2. Untersuchen Sie die in diesen Netzen festgehaltenen Pakete, und suchen Sie nach Beweisen dafür, ob Split Horizon aktiv ist oder nicht.

Übung 4: Verlust einer Route

Material: Topologie aus Übung 1, Wireshark

1. Bei laufendem Wireshark trennen Sie die Verbindung zwischen dem Netzwerk 192.168.3.0 und R2.

2. Welcher Traffic wird daraufhin erzeugt? Wie schnell sind die Pakete zu sehen?
3. Untersuchen Sie den Inhalt der Pakete. Ist irgendetwas an den Informationen für das 192.168.3.0-Netzwerk von Bedeutung?
4. Stellen Sie die Topologie für das nächste Experiment wieder her.

Übung 5: Timer

Material: Topologie aus Übung 1, Wireshark, ein Switch zwischen R1 und R2

1. Beobachten Sie die Häufigkeit, mit der RIP-Pakete von den Routern gesendet werden. Entspricht das dem Timer-Wert aus diesem Kapitel?
2. Bei laufendem Wireshark trennen Sie die Verbindung zwischen dem Netzwerk 192.168.2.0 und R2. Beschreiben Sie aus Sicht von R1 die Unterschiede zwischen diesem und dem vorherigen Experiment.
3. Beobachten Sie die Routingtabelle von R1. Wie lange dauert es, bis der Eintrag für das 192.168.3.0-Netzwerk verschwindet?
4. Gab es aufgrund dieser Unterbrechung Änderungen in den Paketen? Hinweis: Glaubt R2, dass das 192.168.3.0-Netzwerk weg ist?
5. Welche Timer sind mit R1 verknüpft? Praktischer Cisco-Befehl: `show ip protocol`

KAPITEL 6

Open Shortest Path First

In diesem Kapitel:

- Beschreibung des Protokolls
- Link State
- Struktur und grundlegender Betrieb
- Fortgeschrittener Betrieb
- OSPF und IPv6
- Lektüre
- Zusammenfassung
- Fragen
- Antworten
- Laborübungen

Open Shortest Path First (OSPF) und das Routing Information Protocol (RIP) sind beides interne Routingprotokolle, die für Netzwerke innerhalb eines einzelnen autonomen Systems entwickelt wurden. Doch damit endet die Ähnlichkeit auch schon. OSPF verwendet einen völlig anderen Algorithmus und ist ein Link-State-Protokoll, während RIP ein Distanzvektor-Protokoll ist. Im letzten Kapitel haben wir RIP als einfaches, aber zuverlässiges Protokoll kennengelernt. Die Einschränkungen im Hinblick auf die Netzwerk-Größe und die langsame Konvergenzzeit schränken den Einsatz von RIP allerdings ein. Wachsen Netzwerk-Topologien über 15 Hops hinaus oder werden die Topologien komplexer, sind Protokolle wie OSPF nicht nur attraktiver, sondern unabdingbar.

Auch dieses Kapitel baut die Beispiel-Topologien auf Cisco-Geräten auf, um die Ideen, Pakete und den Betrieb von OSPF zu untersuchen. Wie bei den meisten Protokollen ist das Einrichten und der Betrieb von OSPF nur eine Sache von wenigen Befehlen. Allerdings kann die Komplexität von OSPF sehr schnell anwachsen. Für OSPF gibt es eine ausführliche Spezifikation und eine Reihe von RFCs, die das ursprüngliche Protokoll erweitern. Damit das Kapitel übersichtlich bleibt, werden hier nur die am häufigsten genutzten Features diskutiert.

Beschreibung des Protokolls

Die Spezifikation zu Open Shortest Path First ist 1989 erstmals in RFC 1131 erschienen, wurde aber zwei Jahre später durch RFC 1247 ersetzt, das OSPF Version 2 behandelt. Auch diese Version des Protokolls wurde wiederholt aktualisiert. Für den Zweck dieses Kapitels bildet RFC 2328 (und Paket-Captures) die Grundlage unserer Betrachtung. Im Gegensatz zu RIP werden OSPF-Nach-

richten direkt in IP-Paketen gekapselt. Die IP-Protokoll-ID für OSPF ist 89, d.h., das Herausfiltern von UDP-Streams führt nicht wie bei RIP zu Problemen. Die Kapselung und die IP-Protokoll-ID sind in Abbildung 6-1 zu sehen.

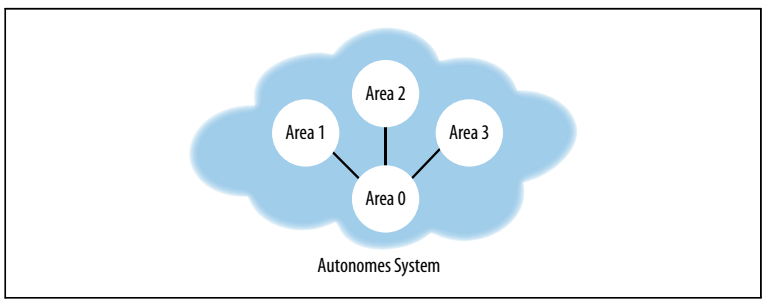
```
Ethernet II, Src: Cisco_da:5a:a0 (00:05:32:da:5a:a0), Dst: IPv4mcast_00:00:05 (01:00:5e:00:00:05)
Internet Protocol, Src: 192.168.2.253 (192.168.2.253), Dst: 224.0.0.5 (224.0.0.5)
  Version: 4
  Header length: 20 bytes
⊞ Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
  Total Length: 64
  Identification: 0x3a50 (14928)
⊞ Flags: 0x00
  Fragment offset: 0
  Time to live: 1
  Protocol: OSPF IGP (89)
⊞ Header checksum: 0xdaaa [correct]
  Source: 192.168.2.253 (192.168.2.253)
  Destination: 224.0.0.5 (224.0.0.5)
Open Shortest Path First
```

Abbildung 6-1 ▲ OSPF-Kapselung und IP-Protokoll-ID

Distanzvektor-Protokolle senden regelmäßige Updates, die die gesamte Routingtabelle oder einen Teil von ihr enthalten. Solche »Ankündigungen« (Advertisements) wurden um Techniken wie Split Horizon, Poisoning, getriggerte Updates und Count-to-Infinity erweitert. Link-State-Protokolle verwenden einen anderen Ansatz, bei dem »Link-State-Advertisements« per Multicast-Flooding über die gesamte Topologie verteilt werden, doch sobald sich die Topologie stabilisiert hat, hören die Updates auf. Stattdessen wird eine Hello-Nachricht gesendet, die anzeigt, dass sich nichts geändert hat. In Hello-Nachrichten findet sich nur wenig nützlicher Inhalt. Router, die OSPF nutzen, entwickeln ein Bild der Topologie und speichern es in der Link-State-Datenbank (Link State Database, LSDB) jedes Routers.

OSPF-Netzwerke organisieren sich um »Areas« (Bereiche) herum. Eine solche Area besteht aus einer Gruppe von Routern, die Traffic für eine Gruppe von Netzwerken weiterleiten. Ein OSPF-AS (autonomes System) ist eine Gruppe von OSPF-Areas. Man kann sich OSPF auch als kleinere Topologien innerhalb der AS-Topologie vorstellen. Router innerhalb einer Area wissen nicht viel über die Welt außerhalb dieser Area. Die grundlegende Idee ist in Abbildung 6-2 dargestellt.

Abbildung 6-2 ▶ Autonomes System mit Areas



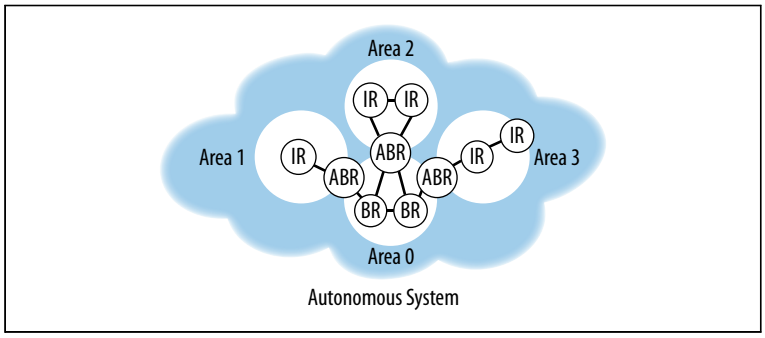
Jede Area betreibt ihre eigene Instanz des Link-State-Protokolls und entwickelt mit der Zeit einen eigenen Baum kürzester Pfade und Routingtabellen für die Router. Areas wird ein numerischer Wert zugewiesen, und dieser kann entweder als einzelne Zahl oder in Punktnotation (wie eine IP-Adresse) dargestellt werden. In den Captures dieses Kapitels wird Area 51 zu 0.0.0.51. Netzwerke jedweder Größe haben einen »Backbone«-Bereich. Dem Backbone wird immer die Area 0 zugeordnet. Areas müssen physisch, aber nicht virtuell zusammenhängen. Das bedeutet, dass eine Area auf beiden Seiten eines virtuellen Links, etwa einer Verbindung zu einem virtuellen privaten Netzwerk, existieren kann. Pakete, die zwischen Nicht-Backbone-Areas fließen, laufen über das Backbone. Der Traffic ist entweder Intra-Area (Quelle und Ziel liegen in einer Area) oder Inter-Area, wobei der Traffic von einer Area zum Backbone und dann zu einer anderen Area läuft. In RFC 2328 heißt es dazu (frei übersetzt):

> Anders betrachtet, kann man sich das Inter-Area-Routing als erzwungene Stern-Konfiguration im autonomen System vorstellen. Das Backbone bildet den Mittelpunkt, und die Nicht-Backbone-Areas sind die Speichen.

Eine OSPF-Topologie umfasst verschiedene Arten von Routern:

- Area Border Router (ABR) – verbinden Areas.
- Interne Router (IR) – Router innerhalb einer Area.
- Backbone-Router (BR) – Router mit einem Interface zum Backbone. Kann ein ABR sein.
- AS Boundary Router (ASBR) – befindet sich an der Grenze (Boundary) des AS und tauscht Informationen mit anderen ASBRs aus. Dieser Router bietet innerhalb der lokalen OSPF-Topologie auch externe Routen an.

Fügt man diese Router in das Diagramm in Abbildung 6-2 ein, sieht es so aus wie in Abbildung 6-3.



◀ Abbildung 6-3
Modifiziertes AS-Diagramm

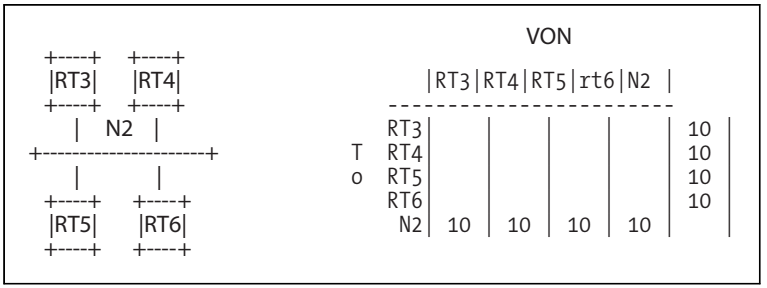
Aus Abbildung 6-3 kann man ersehen, dass die Topologien der einzelnen Areas variieren können. Das Routing innerhalb eines autonomen Systems wird über OSPF-Link-State-Advertisements, kurz LSAs, geregelt. Im Falle einer sogenannten Stub-Area (bei der nur ein Weg nach draußen führt), bietet der ABR typischerweise eine Default-Route an. Darüber hinaus werden extern gewonnene LSAs in Stub-Areas nicht bekannt gemacht, da man sie dort nicht braucht. OSPF unterstützt das Routing an extern gewonnene Ziele (»Blätter« im OSPF-Baum) und CIDR (Classless Inter-Domain Routing). Die extern gewonnenen Routen werden durch das autonome System bekannt gegeben.

Link State

Der *Status* (*state*) beschreibt die Lage des Routers innerhalb der Topologie. Dieser Status besteht aus den verbundenen Interfaces des Routers und den Nachbarn, die über diese Interfaces erreichbar sind. Aus RFC 2328 (frei übersetzt):

> Aus der Link-State-Datenbank konstruiert jeder Router einen Baum der kürztesten Pfade mit sich selbst als Wurzel (Root). Dieser Kürzester-Pfad-Baum (Shortest-Path-Tree) liefert die Route zu jedem Ziel innerhalb des automomen Systems.

Link-State-Protokolle werden auch als Shortest-Path-First-Protokolle (»kürzester Pfad zuerst«) bezeichnet oder basieren zumindest auf Shortest-Path-First-Algorithmen. Wie die Distanzvektor-Protokolle gibt es sie schon seit Jahrzehnten. Im Grunde wird das Netzwerk als Graph betrachtet. Die Router sind die (Eck-)Punkte des Graphen, und die dazwischen liegenden Links werden als Kanten bezeichnet. Die Kanten können einen Router auch mit einem Netzwerk verbinden. Bei OSPF sind die Netzwerke entweder Transit- oder Stub-Netzwerke. Der Traffic in Transit-Netzwerken unterscheidet sich von dem in Stub-(Stummel-)Netzwerken dadurch, dass er »weder lokal generiert noch lokal bestimmt« sein muss. Darüber hinaus können mit einem Netzwerk ein oder mehrere Router verbunden sein, und zwar unabhängig von deren Typ. Die Kosten einer bestimmten Route oder eines bestimmten Links werden mit der Ausgabeseite des Router-Interfaces verknüpft. Niedrigere Kosten sind besser. Aus dieser Gruppe von Netzwerktypen, Kosten und IP-basierten Verbindungen kann ein Graph gezeichnet werden, der die Konnektivität darstellt. Ein sehr einfaches Beispiel ist in Abbildung 6-4 zu sehen. Dieses Diagramm wurde aus RFC 2328 übernommen und mit beispielhaften Kosten für langsame Ausgabe-Links versehen.



◀ **Abbildung 6-4**
Beispiel-Graph

Das Diagramm links zeigt vier Router, die mit dem Netzwerk N2 verbunden sind. Rechts sehen Sie eine grafische Darstellung der Netzwerk-Konnektivität und die Geschwindigkeit der Verbindungen. Die Kosten (Metrik) für ein bestimmtes Ziel bildet eine einzelne »dimensionslose« Metrik, die einen Faktor der Netzwerkbedingungen darstellt. Das ist anders als bei RIP, wo ein einfacher Hop-Zähler als Metrik verwendet wird.

Während die Router LSAs senden (Link State Advertisements), kann eine LSDB (Link State-Datenbank) aufgebaut werden, die Informationen zur Nachbarschaft des Routers enthält. Die LSAs fluten das gesamte Netzwerk, und jeder Router fügt immer weitere Details hinzu. *Ohne unterschiedliche Areas würde jeder Router letztlich die Topologie des gesamten autonomen Systems kennen, und alle würden die gleiche LSDB aufweisen*. Mithilfe der LSDB entwickeln die Router eigene Bäume mit den kürzesten Pfaden zu jedem Ziel. Das bedeutet, dass die LSDBs zwar gleich sein können, die daraus resultierenden Bäume aber nicht. Die Konnektivität zu Zielen, die außerhalb des OSPF-AS liegen, kann einfach erreicht werden, indem man die Routen intern anbietet. Die Metriken für die Advertisements nennt man Typ I und Typ II. Typ I-Metriken wandeln die externen Kosten und die OSPF-Kosten um und addieren diese.

- Typ I = Kosten zum anbietenden Router + Kosten zum externen Ziel
- Typ II wandelt nichts um, sondern bietet nur die geringsten externen Kosten zum Ziel an.

Sobald eine Gruppe zusammenhängender Netzwerke in Areas aufgeteilt wurde, werden separate Kopien des Link-State-Algorithmus ausgeführt. Dadurch entwickeln sich eigene LSDBs und Bäume kürzester Pfade. Wie bereits erwähnt wurde, wissen die Router innerhalb einer Area nicht viel über das gesamte AS oder über andere Areas. Bei der Topologie aus Abbildung 6-2 gibt es mindestens vier verschiedene Instanzen des Link-State-Algorithmus. Das reduziert den gesamten Routing-Traffic, da einzelne Router nicht

alles überall anbieten müssen. Die Router sind auf einen Area Border Router (ABR) angewiesen, der Traffic an Ziele weiterleitet, die sich außerhalb der Area befinden.

OSPF-Router kommunizieren recht viel, um die LSDB aufzubauen. Im nächsten Abschnitt wollen wir eine kleine Topologie nutzen, um die verwendeten Nachrichten und Felder zu erläutern, die dazu notwendig sind.

Struktur und grundlegender Betrieb

Dieser Abschnitt behandelt den Inhalt und die Struktur von fünf OSPF-Nachrichten. Die kleine Topologie aus Abbildung 6-5 bildet die Umgebung für die Nachrichten. Anfangs ist die Topologie so konfiguriert, wie auch die anderen Topologien in diesem Buch. IP-Adressen werden den Interfaces zugeordnet, und die Schnittstellen werden aktiviert. Sobald die Verkabelung erledigt ist, enthalten die Routingtabellen von R1 und R2 die Einträge aus Tabelle 6-1.

Abbildung 6-5 ▶
Kleine OSPF-Topologie

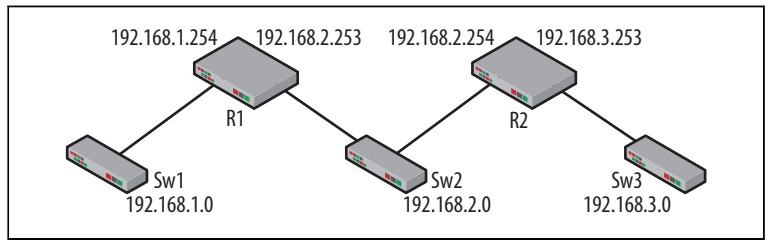


Tabelle 6-1: Anfängliche Routingtabellen

R1	R2
C 192.168.1.0/24	C 192.168.2.0/24
C 192.168.2.0/24	C 192.168.3.0/24

Nachdem die Topologie steht und die Routingtabellen aufgebaut sind, müssen die Router für den Betrieb von OSPF konfiguriert werden. Dazu werden die folgenden Befehle auf R1 und R2 eingegeben:

```
R1
    router ospf 10
    network 192.168.1.0 255.255.255.0 area 51
    network 192.168.2.0 255.255.255.0 area 51
R2
    router ospf 20
    network 192.168.2.0 255.255.255.0 area 51
    network 192.168.3.0 255.255.255.0 area 51
```

Der erste Befehl stößt den Prozess an und vergibt die Prozessnummer für den Router. Die Prozess-ID identifiziert die OSPF-Instanz. Auf einem Router kann mehr als ein OSPF-Prozess laufen. Router innerhalb des gleichen OSPF-AS müssen nicht die gleiche Prozess-ID verwenden. Der `network`-Befehl gibt das anzubietende Netzwerk an und weist dem Netzwerk eine Area zu.

Hello

Während die Befehle eingegeben und die Netzwerke in die Advertisement-Liste eingefügt werden, beginnen die Router (von links nach rechts) damit, die erste OSPF-Nachricht zu senden: Hello. Sobald R1 zum Beispiel den Befehl `network 192.168.1.0 255.255.255.0 area 51` erhält, sendet er sofort eine »Hello«-Nachricht an das Netzwerk. Ein Beispiel für eine solche Hello-Nachricht ist in Abbildung 6-6 zu sehen.

```
Ethernet II, Src: Cisco_28:02:81 (00:05:5e:28:02:81), Dst: IPv4mcast_00:00:05 (01:00:5e:00:00:05)
Internet Protocol, Src: 192.168.2.254 (192.168.2.254), Dst: 224.0.0.5 (224.0.0.5)
Open Shortest Path First
⊟ OSPF Header
    OSPF Version: 2
    Message Type: Hello Packet (1)
    Packet Length: 44
    Source OSPF Router: 192.168.3.253 (192.168.3.253)
    Area ID: 0.0.0.51
    Packet Checksum: 0x37c6 [correct]
    Auth Type: Null
    Auth Data (none)
⊟ OSPF Hello Packet
    Network Mask: 255.255.255.0
    Hello Interval: 10 seconds
  ⊟ Options: 0x02 (E)
      0... .... = DN: DN-bit is NOT set
      .0.. .... = O: O-bit is NOT set
      ..0. .... = DC: Demand Circuits are NOT supported
      ...0 .... = L: The packet does NOT contain LLS data block
      .... 0... = NP: NSSA is NOT supported
      .... .0.. = MC: NOT Multicast Capable
      .... ..1. = E: External Routing Capability
      .... ...0 = MT: NO Multi-Topology Routing
    Router Priority: 1
    Router Dead Interval: 40 seconds
    Designated Router: 0.0.0.0
    Backup Designated Router: 0.0.0.0
```

▲ **Abbildung 6-6**
OSPF-Hello

Die Hello-Nachricht erlaubt es dem Router, sich selbst anzubieten und etwas über andere OSPF-Nachbarn zu erfahren. Die Idee eines »Nachbarn« ist wichtig für OSPF. Da der Router etwas über seine Umgebung lernt, verwenden Hello-Nachrichten die OSPF-spezifische Multicast-Adresse 224.0.0.5. Man spricht bei einer Gruppe von Routern, die etwas auf diese Weise lernen, vom Betrieb eines OSPF-Broadcast-Netzwerks. Sobald die Nachbarn bekannt sind, werden Hello-Nachrichten zur Pflege der Beziehungen verwendet. Router stellen eine »Nähe« zu den benachbarten Routern her. Im

gleichen Netzwerk miteinander verbundene Router sind Nachbarn. Allerdings tauschen nur angrenzende Router Routinginformationen miteinander aus. Diese Nachbarschaft ist erwünscht, wenn die Router Hello-Nachrichten austauschen und die OSPF-Optionen kompatibel sind. Einmal hergestellt, tauchen die Nachbar-Adressen in den Hello-Nachrichten auf, Datenbankinformationen können ausgetauscht und auf dem neuesten Stand gehalten werden.

Die OSPF-Hello-Nachricht ist bei einer einfachen Konfiguration ganz simpel:

Version
: Dieses 1-Byte-Feld hat nahezu immer den Wert 2.

Nachrichten-Typ
: Dieses 1-Byte-Feld enthält den numerischen Code der Nachricht. Eine Hello-Nachricht hat den Code 1.

Paketlänge
: Die Länge des OSPF-Pakets.

Quell-OSPF-Router
: Dies ist ein konfigurierbarer Wert für die IP-Adresse des Quell-Routers.

Area-ID
: Diese 4 Bytes geben die für dieses Netzwerk konfigurierte Area an.

Paket-Prüfsumme
: Die 16-Bit-Einerkomplement-Prüfsumme des OSPF-Pakets

Authentifizierungstyp und -daten
: Das RFC verlangt, dass OSPF-Pakete entsprechend dem in diesem Feld stehenden Typ authentifiziert werden müssen. Fehlt die Konfiguration, wird (wie es oben zu sehen ist) mit Nullen aufgefüllt.

Netzwerkmaske
: Maske des in der Routerkonfiguration angegebenen Netzwerks.

Hello-Interval
: Häufigkeit, mit der Hello-Nachrichten generiert werden; 2 Bytes.

Optionen
: Dieses 1-Byte-Feld beschreibt die Betriebscharakteristika des OSPF-Routers. Bei einer Grundkonfiguration ändern sich die Optionen größtenteils nicht. Die Grundidee lautet, dass die Optionen zwischen den Nachbarn übereinstimmen sollten. Die kritische Option ist das E-Bit. Dieses Feld ist in verschiedenen Pakettypen enthalten und wird üblicherweise genutzt, um

anzuzeigen, dass Routing-Advertisements im lokalen Netzwerk übertragen werden. Fünf Bits dieses Feldes wurden in RFC 2328 definiert, weitere wurden in späteren RFCs hinzugefügt.

DN
: Zeigt an, dass diese Route bei OSPF-Berechnungen per RFC 2547 nicht verwendet werden soll.

O
: Unterstützung für opake LSAs (wie in RFC 2370 definiert)

DC
: Zeigt die Unterstützung von Demand Circuits (wie in RFC 3883 (1793) definiert) an.

L
: Zeigt die Unterstützung der Link-Local-Signalisierung für zusätzliche Routendaten an (siehe RFC 5613).

NP
: Handling von sogenannten Typ-7 Not-So-Stubby-Areas-(NSSA-)LSAs (siehe RFC 1587).

MC
: Zeigt an, ob Multicast-Forwarding nach RFC 1584 unterstützt wird.

E
: Beschreibt die Flooding-Methode für LSAs.

MT
: Bietet Unterstützung für verschiedene Topologie-Arten oder Traffic wie QoS.

Router-Priorität
: Ein 1-Byte-Feld, das zur Wahl des designierten Routers für das Netzwerk verwendet wird.

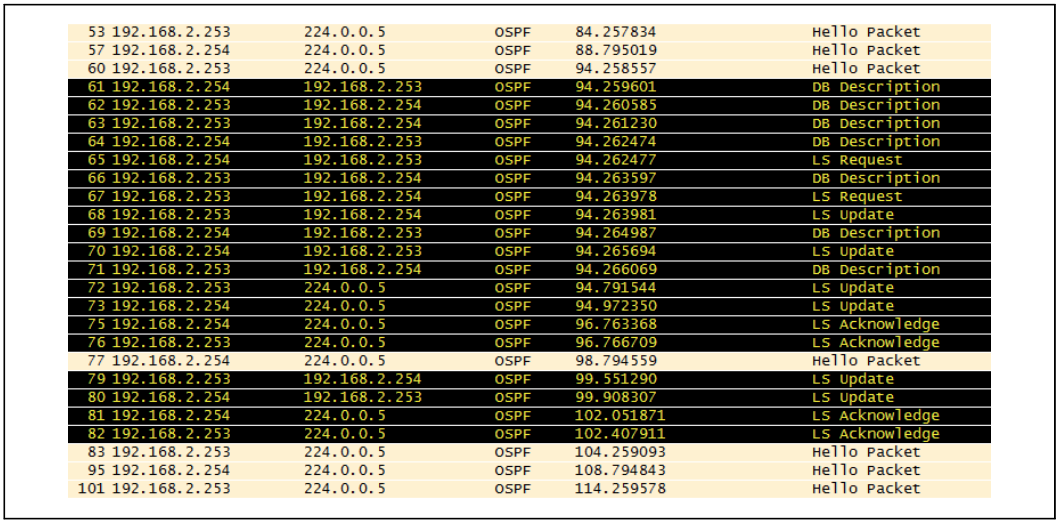
Router-Tot-Intervall (Router Dead Interval)
: Gibt an, wie oft ein Router gehört werden muss, damit er immer noch als aktiver Nachbar betrachtet wird.

Designierter Router
: Router, der den Status des Netzwerks anbietet

Backup für designierten Router
: Übernimmt die Rolle des designierten Routers, wenn dieser ausfällt.

Während die Befehle eingegeben werden, tauschen R1 und R2 Hello-Nachrichten aus und richten sich als Nachbarn ein. Sobald das geschehen ist, tauschen sie Routinginformationen per Link State-

Updates aus. Die Kommunikation ist in Abbildung 6-7 zu sehen. Sie umfasst den Standard-Satz der OSPF-Nachrichtentypen.

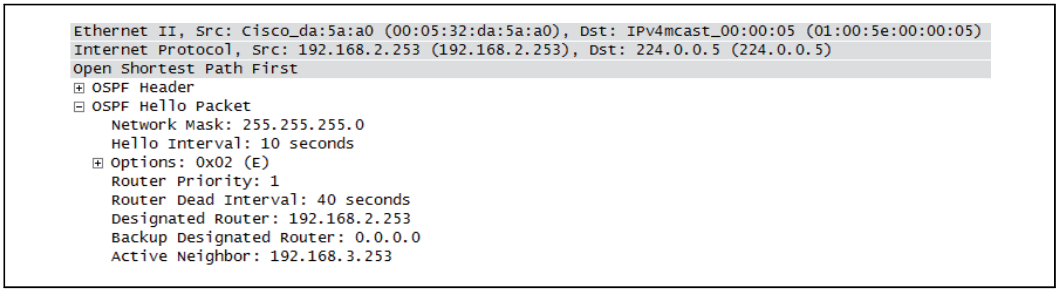


53	192.168.2.253	224.0.0.5	OSPF	84.257834	Hello Packet
57	192.168.2.254	224.0.0.5	OSPF	88.795019	Hello Packet
60	192.168.2.253	224.0.0.5	OSPF	94.258557	Hello Packet
61	192.168.2.254	192.168.2.253	OSPF	94.259601	DB Description
62	192.168.2.253	192.168.2.254	OSPF	94.260585	DB Description
63	192.168.2.253	192.168.2.254	OSPF	94.261230	DB Description
64	192.168.2.254	192.168.2.253	OSPF	94.262474	DB Description
65	192.168.2.254	192.168.2.253	OSPF	94.262477	LS Request
66	192.168.2.253	192.168.2.254	OSPF	94.263597	DB Description
67	192.168.2.253	192.168.2.254	OSPF	94.263978	LS Request
68	192.168.2.253	192.168.2.254	OSPF	94.263981	LS Update
69	192.168.2.254	192.168.2.253	OSPF	94.264987	DB Description
70	192.168.2.254	192.168.2.253	OSPF	94.265694	LS Update
71	192.168.2.253	192.168.2.254	OSPF	94.266069	DB Description
72	192.168.2.253	224.0.0.5	OSPF	94.791544	LS Update
73	192.168.2.254	224.0.0.5	OSPF	94.972350	LS Update
75	192.168.2.254	224.0.0.5	OSPF	96.763368	LS Acknowledge
76	192.168.2.253	224.0.0.5	OSPF	96.766709	LS Acknowledge
77	192.168.2.254	224.0.0.5	OSPF	98.794559	Hello Packet
79	192.168.2.253	192.168.2.254	OSPF	99.551290	LS Update
80	192.168.2.254	192.168.2.253	OSPF	99.908307	LS Update
81	192.168.2.254	224.0.0.5	OSPF	102.051871	LS Acknowledge
82	192.168.2.253	224.0.0.5	OSPF	102.407911	LS Acknowledge
83	192.168.2.253	224.0.0.5	OSPF	104.259093	Hello Packet
95	192.168.2.254	224.0.0.5	OSPF	108.794843	Hello Packet
101	192.168.2.253	224.0.0.5	OSPF	114.259578	Hello Packet

Abbildung 6-7 ▲
OSPF-Kommunikation zwischen R1 und R2

Beschreibung der DB

Die Datenbankbeschreibungspakete (OSPF Typ 2) werden beim Aufbau nachbarschaftlicher Beziehungen ausgetauscht. Sieht man sich eine der letzten Hello-Nachrichten vor den DB-Beschreibungen an (Paket 60), sieht man ein zusätzliches Feld – den aktiven Nachbarn (Active Neighbor). Das Hello des aktiven Nachbarn ist in Abbildung 6-8 zu sehen und die DB-Beschreibung in Abbildung 6-9.



```
Ethernet II, Src: Cisco_da:5a:a0 (00:05:32:da:5a:a0), Dst: IPv4mcast_00:00:05 (01:00:5e:00:00:05)
Internet Protocol, Src: 192.168.2.253 (192.168.2.253), Dst: 224.0.0.5 (224.0.0.5)
Open Shortest Path First
⊞ OSPF Header
⊟ OSPF Hello Packet
    Network Mask: 255.255.255.0
    Hello Interval: 10 seconds
  ⊞ Options: 0x02 (E)
    Router Priority: 1
    Router Dead Interval: 40 seconds
    Designated Router: 192.168.2.253
    Backup Designated Router: 0.0.0.0
    Active Neighbor: 192.168.3.253
```

Abbildung 6-8 ▲
Aktiver-Nachbar-Hello

An diesem Punkt werden ein oder mehrere DB-Beschreibungspakete zwischen den Routern gesendet. Die Kommunikation wird durch den designierten Router kontrolliert und verwendet eine Poll/Response-Prozedur, bei der ein Router als Master agiert. Dem Austausch der DB-Beschreibung kann man folgen, indem man ne-

ben den DD-Sequenznummern die R-, I-, M- und MS-Bits untersucht.

R Resynchronisationsbit – es erlaubt herstellerspezifische Implementierungen zur Neuberechnung der Link-State-Datenbank (LSDB) ohne Topologieänderung. Beschrieben in RFC 4811.

I Initialisierungsbit – zeigt den Start des Austauschs an.

M Mehr (More) – zeigt an, ob die DB-Beschreibung abgeschlossen ist.

MS Master/Slave – gibt den Router an, der den Austausch kontrolliert. Beachten Sie, dass das DB-Beschreibungspaket in Abbildung 6-9 von 192.168.2.254 (R2) stammt und das dass Master-Bit gesetzt ist.

DD Die Sequenznummer für eine bestimmte Nachricht – Bei gesetztem Init-Flag ist dies das erste Paket. Die DD-Werte erhöhen sich, bis die DB-Beschreibung komplett ist.

Den letzten Abschnitt der Pakete nimmt der LSA-Header ein, der Informationen über die Link-State-Datenbank (LSDB) enthält.

LSA-Header

In den RFCs sind unterschiedliche LSA-Typen definiert. Von diesen sind die *Router-LSAs* (die von allen Routern verwendet werden) und die *Netzwerk-LSAs* (designierter Router) die gängigsten. Zu den anderen Typen gehören »summary« und »external«. Die LSA-Header können sich basierend auf dem LSA-Typ ändern. In Abbildung 6-9 ist der LSA-Typ ein Router-LSA. Diese Nachricht ist Teil der Konversation zwischen R1 und R2. Andere LSAs werden per Multicast über die OSPF-Multicast-Adressen durch die Topologie geflutet. Zusammen genommen bilden die LSAs die Link-State-Datenbank (LSDB). Sobald die Router über die LSDB verfügen, können Sie den Baum mit den kürzsten Pfaden aufbauen (oder neu aufbauen). Erinnern Sie sich daran, dass jeder Router sich selbst als die Wurzel des Baums sieht. Dieser Baum führt dann zur Routingtabelle.

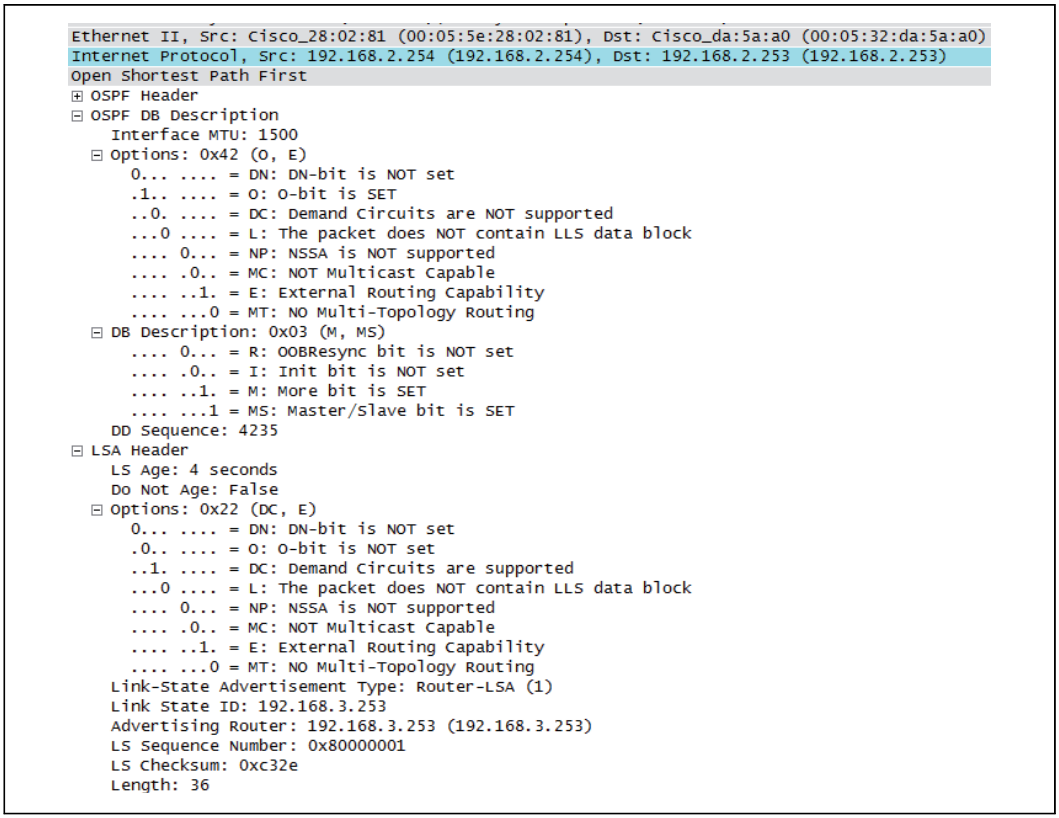
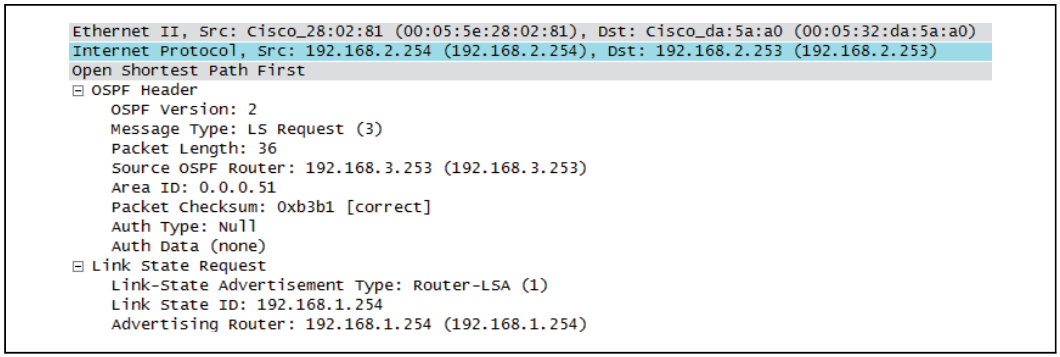


Abbildung 6-9 ▲
DB-Beschreibungsnachricht

Die Optionsfelder stellen die Fähigkeiten des Netzwerks dar. Die Link-State-ID ist entweder die IP-Adresse des designierten Routers oder des Quell-Interfaces. Die LS-Sequenzummer hilft bei der Erkennung von Duplikaten. Die Prüfsumme ist für das jeweilige LSA, klammert aber das LS-Alter aus.



▲ **Abbildung 6-10**
LS-Request-Nachricht

Link-State-Request

Ein LS-Request ist in Abbildung 6-10 zu sehen. Das ist der dritte OSPF-Nachrichtentyp. Als einfachste OSPF-Nachricht enthält sie keinerlei Informationen, die wir nicht bereits erörtert hätten. Sie wird verwendet, um Informationen anzufordern, wenn der Router seine LSDB aktualisieren möchte. In diesem Fall hat der Router noch keine Informationen in der LSDB, versteht aber die Struktur und die Optionen der Routingdomäne. Beachten Sie, dass die Adressierung zwischen den Nachbarn läuft.

Link-State-Update

Das LS-Update in Abbildung 6-11 ist die Response auf einen LS-Request. Das LS-Update ist der vierte Nachrichtentyp für OSPF. Die Aufgabe der Pakete besteht darin, LSAs durch die gesamte Topologie zu fluten. LS-Updates leiten Informationen einen Hop weiter. Auf diese Weise werden die Informationen schließlich überall hin propagiert.

Das Paket in Abbildung 6-11 ist ein Beispiel dafür, wie Wireshark versucht, bei der Entschlüsselung des Pakets behilflich zu sein. Allerdings kann das recht verwirrend sein. Unter jedem Link-Eintrag erscheinen wesentlich mehr Informationen, als im Paket tatsächlich enthalten sind. Um deutlich zu machen, was tatsächlich gesendet wird, habe ich die hexadezimalen Werte unter die Pakete und Beispiele in den Feldbeschreibungen aufgenommen. Jeder LSA-Eintrag ist nur 12 Byte lang und beginnt mit der Netzwerkadresse.

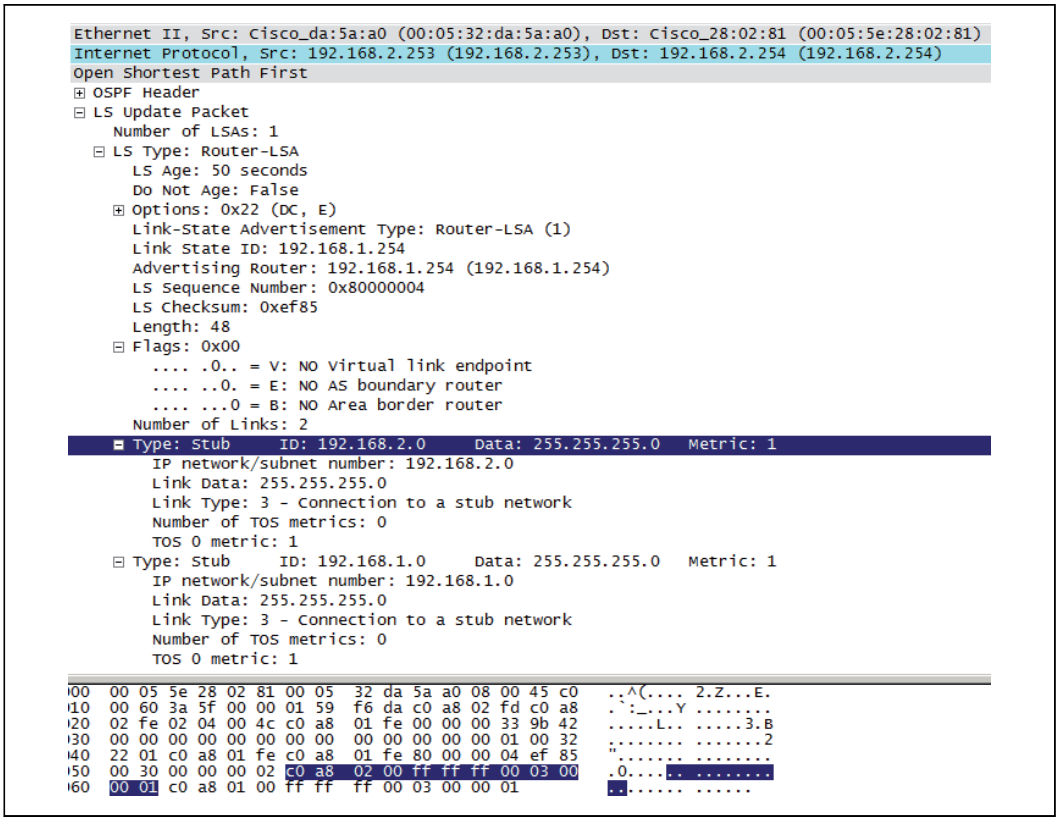


Abbildung 6-11 ▲
LS-Update-Nachricht

Hier die LSA-Konstruktion, beginnend mit dem Flags-Feld:

V

Zeigt an, dass der Router einen Endpunkt für einen virtuellen Link bildet. Wert in Abbildung 6-11: 0

E

Extern – der Router ist ein AS-Boundary-Router. Wert in Abbildung 6-11: 0

B

Der Router ist ein Area-Border-Router. Wert in Abbildung 6-11: 0

Anzahl der Links

Zahl der insgesamt vom Router verarbeiteten Links. Bei dieser Topologie haben beide Router jeweils 2.

Link-ID

Dieser Wert hängt vom Typ des Links ab (siehe Tabelle 6-2). In Abbildung 6-11 entspricht er der Netzwerk-ID 192.168.2.0. Hex: c0 a8 02 00

Tabelle 6-2: OSPF LSA-Link-IDs

Type	Link ID
1	Router-ID des benachbarten Routers
2	IP-Adresse des designierten Routers
3	IP-Netzwerk oder -Subnetzwerk
4	Router-ID des benachbarten Routers

Daten
: Hängt ebenfalls vom Linktyp ab. In diesem Fall hat er die Netzwerkmaske 255.255.255.0. Hex: ff ff ff 00

Linktyp
: Wie in Tabelle 6-3 zu sehen ist, sind vier Linktypen definiert. Bei diesem Paket glaubt R1 anfänglich, dass beide Netzwerke Stub-Netzwerke sind, d.h., dass es keinen anderen Weg hinaus gibt. Daher verwenden beide den Wert 3 und verwenden das IP-Netzwerk als Link-ID.

Tabelle 6-3: OSPF LSA-Linktypen

1	Punkt-zu-Punkt
2	Verbindung zu einem Transit-Netzwerk
3	Verbindung zu einem Stub-Netzwerk
4	Virtueller Link

Anzahl der ToS-Metriken
: »Type of Service«, kurz ToS, ist eine Prioritäts- oder Qualitätseinstellung, die die Behandlung von Paketen beschreibt. IP-ToS-Werte werden nur selten genutzt und wurden größtenteils durch »Differentiated Services« ersetzt. Wurden keine Metriken festgelegt, enthält das Feld (wie in diesem Fall) den Wert 0.

Metrik
: RFC 2328 beschreibt die Metrik einfach als Kosten für die Nutzung dieses Links, auch wenn die Geschwindigkeit des Links ein wichtiger Faktor ist.

Die IP-Adressierung in Abbildung 6-11 stammt aus der Kommunikation zwischen R1 und R2. Abbildung 6-12 zeigt ein weiteres LS-Update der gleichen Kommunikation (Paket 72), doch man kann erkennen, dass R1 zum designierten Router für dieses Segment gekürt wurde.

```
Ethernet II, Src: Cisco_da:5a:a0 (00:05:32:da:5a:a0), Dst: IPv4mcast_00:00:05 (01:00:5e:00:00:05)
Internet Protocol, Src: 192.168.2.253 (192.168.2.253), Dst: 224.0.0.5 (224.0.0.5)
Open Shortest Path First
⊞ OSPF Header
⊟ LS Update Packet
    Number of LSAs: 2
  ⊟ LS Type: Router-LSA
      LS Age: 1 seconds
      Do Not Age: False
    ⊞ Options: 0x22 (DC, E)
      Link-State Advertisement Type: Router-LSA (1)
      Link State ID: 192.168.1.254
      Advertising Router: 192.168.1.254 (192.168.1.254)
      LS Sequence Number: 0x80000005
      LS Checksum: 0xf815
      Length: 48
    ⊞ Flags: 0x00
      Number of Links: 2
    ⊟ Type: Transit  ID: 192.168.2.253   Data: 192.168.2.253   Metric: 1
        IP address of Designated Router: 192.168.2.253
        Link Data: 192.168.2.253
        Link Type: 2 - Connection to a transit network
        Number of TOS metrics: 0
        TOS 0 metric: 1
    ⊟ Type: Stub     ID: 192.168.1.0     Data: 255.255.255.0   Metric: 1
        IP network/subnet number: 192.168.1.0
        Link Data: 255.255.255.0
        Link Type: 3 - Connection to a stub network
        Number of TOS metrics: 0
        TOS 0 metric: 1
  ⊟ LS Type: Network-LSA
      LS Age: 1 seconds
      Do Not Age: False
    ⊞ Options: 0x22 (DC, E)
      Link-State Advertisement Type: Network-LSA (2)
      Link State ID: 192.168.2.253
      Advertising Router: 192.168.1.254 (192.168.1.254)
      LS Sequence Number: 0x80000001
      LS Checksum: 0xa3ef
      Length: 32
      Netmask: 255.255.255.0
      Attached Router: 192.168.1.254
      Attached Router: 192.168.3.253
```

Abbildung 6-12 ▲
LS-Update vom designierten Router

Es gibt weitere signifikante Änderungen in diesem LS-Update. Einer der Links wurde aufgrund von Advertisements anderer Router dieses Netzwerks als *Transit-Netzwerk* identifiziert. Daher wurden auch der Linktyp, die ID und die Datenfelder entsprechend aktualisiert. Das Netzwerk 192.168.1.0 ist auch weiterhin ein Stub-Netzwerk. Und schließlich, da dieses Update von einem designierten Router stammt, ist dieses LSA ein Netzwerk-LSA. Die Ziel-IP-Adressierung ist zur »SPF-alle-Router-Adresse« zurückgekehrt.

Link-State-ACK

Der Betriebssicherheit zuliebe wird jedes LS-Update-Paket von den Routern im gleichen Netzwerk mit einer LS-ACK-Nachricht vom Typ 5 bestätigt (»acknowledge«). Die LS-ACK enthält die LS-Sequenznummer. Mit einer einzelnen LS-ACK-Nachricht kann mehr als ein LS-Update bestätigt werden.

```
Ethernet II, Src: Cisco_28:02:81 (00:05:5e:28:02:81), Dst: IPv4mcast_00:00:05 (01:00:5e:00:00:05)
Internet Protocol, Src: 192.168.2.254 (192.168.2.254), Dst: 224.0.0.5 (224.0.0.5)
Open Shortest Path First
⊞ OSPF Header
⊟ LSA Header
    LS Age: 50 seconds
    Do Not Age: False
  ⊟ Options: 0x22 (DC, E)
      0... .... = DN: DN-bit is NOT set
      .0.. .... = O: O-bit is NOT set
      ..1. .... = DC: Demand Circuits are supported
      ...0 .... = L: The packet does NOT contain LLS data block
      .... 0... = NP: NSSA is NOT supported
      .... .0.. = MC: NOT Multicast Capable
      .... ..1. = E: External Routing Capability
      .... ...0 = MT: NO Multi-Topology Routing
    Link-State Advertisement Type: Router-LSA (1)
    Link State ID: 192.168.1.254
    Advertising Router: 192.168.1.254 (192.168.1.254)
    LS Sequence Number: 0x80000004
    LS Checksum: 0xef85
    Length: 48
⊟ LSA Header
    LS Age: 1 seconds
    Do Not Age: False
  ⊟ Options: 0x22 (DC, E)
      0... .... = DN: DN-bit is NOT set
      .0.. .... = O: O-bit is NOT set
      ..1. .... = DC: Demand Circuits are supported
      ...0 .... = L: The packet does NOT contain LLS data block
      .... 0... = NP: NSSA is NOT supported
      .... .0.. = MC: NOT Multicast Capable
      .... ..1. = E: External Routing Capability
      .... ...0 = MT: NO Multi-Topology Routing
    Link-State Advertisement Type: Network-LSA (2)
    Link State ID: 192.168.2.253
    Advertising Router: 192.168.1.254 (192.168.1.254)
    LS Sequence Number: 0x80000001
    LS Checksum: 0xa3ef
    Length: 32
```

▲ **Abbildung 6-13**
LS-ACK

Sobald der Informationsaustausch abgeschlossen ist, kehren die Router zu ihrem normalen (stabilen) Betrieb mit Hello-Nachrichten zurück. Zu diesem Zeitpunkt sind die Routingtabellen unserer Topologie vollständig gefüllt. Die Routingtabellen für R1 und R2 sind in Abbildung 6-14 zu sehen.

▼ **Abbildung 6-14**
Aktualisierte Routingtabellen

```
Gateway of last resort is not set

C    192.168.1.0/24 is directly connected, FastEthernet0/1
C    192.168.2.0/24 is directly connected, FastEthernet0/0
O    192.168.3.0/24 [110/2] via 192.168.2.254, 03:43:11, FastEthernet0/0
R1#

Gateway of last resort is not set

O    192.168.1.0/24 [110/2] via 192.168.2.253, 03:44:38, FastEthernet0/1
C    192.168.2.0/24 is directly connected, FastEthernet0/1
C    192.168.3.0/24 is directly connected, FastEthernet0/0
R2#
```

Diese Routingtabellen spiegeln die Topologie in Abbildung 6-4 wider, bei der R1 und R2 direkt mit zwei Netzwerken verbunden sind, aber keiner von beiden mit allen Zielen direkt verbunden ist. OSPF wurde gestartet und der Paketaustausch aus Abbildung 6-6 abgeschlossen. Die Routingtabellen zeigen die zusätzlichen dynamischen Einträge, die über OSPF gelernt wurden. Jeder Eintrag enthält das Zielnetzwerk und Informationen zum nächsten Hop für dieses Netzwerk. Die Tabelle gibt auch die OSPF-spezifische administrative Distanz und die Metrik in eckigen Klammern an: [110/2]. Die administrative Distanz ist eine Evaluierung der Qualität der OSPF-Informationen relativ zu anderen Routingprotokollen. OSPF hat eine administrative Distanz von 110. RIP hat eine administrative Distanz von 120. Die Metrik wird aus einer Reihe von Faktoren abgeleitet und ist nicht einfach ein Hop-Zähler.

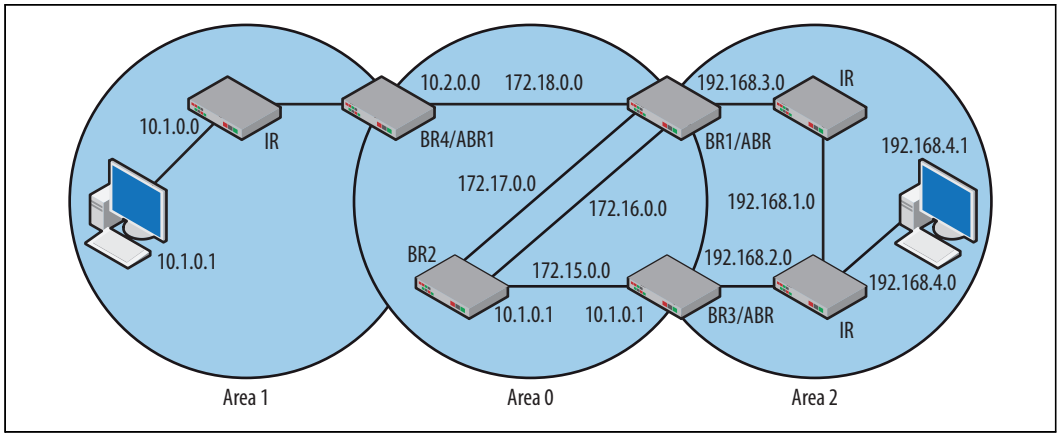
Timer

OSPF definiert zwei allgemeine Arten von Timern: Einzelauslösung (»Single Shot«) und Intervall. Single-Shot-Timer werden für Events wie LS-Updates und Routingänderungen verwendet. Intervall-Timer decken solche Dinge wie Hello-Nachrichten ab. Die Hello-Pakete enthalten einige dieser Werte:

- Hello-Intervall – Es gibt an, wie oft OSPF-Hello-Nachrichten gesendet werden. Der aktuelle Wert liegt bei 10 Sekunden.
- Router-Tot-Intervall (Router Dead Interval) – Läuft dieser Timer aus, verlässt ein Router den Wartezustand. Er beschreibt auch die Zeitspanne, nach der ein Router als »tot« betrachtet wird (d.h., es werden keine Hello-Nachrichten mehr empfangen). Der aktuelle Wert liegt bei 40 Sekunden. Nach dieser Zeitspanne werden Routen, die von diesem Nachbarn stammen, aus der Routingtabelle entfernt.

Fortgeschrittener Betrieb

Die erste in diesem Kapitel verwendete Topologie ist sehr einfach und besteht aus nur einer Area. Wenn Sie das Buch bis hierhin gelesen haben, werden Sie wohl bemerkt haben, dass es eigentlich sinnlos ist, bei einem so kleinen Netzwerk ein Routingprotokoll zu verwenden. Nachdem wir die Paket-Typen und Felder diskutiert und die Terminologie-Hürde genommen haben, wollen wir uns nun einigen komplexeren Ideen und Abbildung 6-15 zuwenden.



▲ **Abbildung 6-15**
OSPF-Topologie mit drei Areas

Abbildung 6-15 zeigt eine weitere OSPF-Topologie, die auf den ersten Blick etwas komplex wirken mag. Doch ein genauerer Blick zeigt, dass es nur drei Areas gibt: 1, 2 und das Backbone (0). In zwei dieser Areas (0 und 2) sind Schleifen integriert, um Redundanz zu schaffen. In den Backbone-Bereichen eines Netzwerks sind Netzwerk-Schleifen eine typische Reaktion auf Flexibilitätsprobleme. Außer auf den Hosts läuft auf allen abgebildeten Geräten OSPF.

Area 1 hat eine sehr einfache Topologie und wird als Stub-Area betrachtet. Die 10.0.0.0-Netzwerke sind auf Area 1 beschränkt. Auf der rechten Seite enthält Area 2 alle 192.168.0.0-Subnetze und weist eine größere Zahl von Verbindungen auf. In diesem Fall führt mehr als ein Weg aus der Area heraus, und der mögliche Traffic könnte einfach durchlaufen, anstatt ein Ziel in einem der beiden 192.168.0.0-Subnetze anzusteuern. Das macht Area 2 zu einer Transit-Area. Area 1 und Area 2 besitzen beide Area-Border-Router (ABRs), die sie mit der zentralen Backbone-Area verbinden. Im Falle von Area 2 dienen zwei Router als ABRs. Die Backbone-Router (BR1–4) liegen in der Backbone-Area, doch drei von ihnen fungieren auch als ABRs.

Abbildung 6-16 zeigt die Routingtabelle von BR4, nachdem die Topologie konvergiert ist. Es gibt insgesamt 10 Netzwerke in dieser Topologie. Davon sind zwei (10.2.0.0 und 172.18.0.0) direkt mit BR4 verbunden, was sich sowohl im Topologie-Diagramm als auch in der Routingtabelle widerspiegelt.

```
O    172.15.0.0/16 [110/3] via 172.18.0.253, 00:19:55, FastEthernet0/0
O    172.17.0.0/16 [110/2] via 172.18.0.253, 00:19:55, FastEthernet0/0
O    172.16.0.0/16 [110/2] via 172.18.0.253, 00:19:55, FastEthernet0/0
C    172.18.0.0/16 is directly connected, FastEthernet0/0
O IA 192.168.4.0/24 [110/4] via 172.18.0.253, 00:19:55, FastEthernet0/0
     10.0.0.0/16 is subnetted, 2 subnets
C       10.2.0.0 is directly connected, FastEthernet0/1
O       10.1.0.0 [110/2] via 10.2.0.254, 00:08:38, FastEthernet0/1
O IA 192.168.1.0/24 [110/3] via 172.18.0.253, 00:19:56, FastEthernet0/0
O IA 192.168.2.0/24 [110/4] via 172.18.0.253, 00:08:39, FastEthernet0/0
O IA 192.168.3.0/24 [110/2] via 172.18.0.253, 00:19:56, FastEthernet0/0
```

Abbildung 6-16 ▲ Routingtabelle einer großen Topologie

BR4 ist Mitglied zweier unterschiedlicher OSPF-Areas. Zur Konfiguration dieses Routers wurden die folgenden Befehle ausgeführt:

```
interface FastEthernet0/0
    ip address 172.18.0.254 255.255.0.0
interface FastEthernet0/1
    ip address 10.2.0.253 255.255.0.0
router ospf 10
    log-adjacency-changes
    network 10.2.0.0 0.0.255.255 area 1
    network 172.18.0.0 0.0.255.255 area 0
```

Wie wir bereits früher in diesem Kapitel gesehen haben, führen die LSA-Nachrichten zwischen den Routern nicht nur Netzwerk-, sondern auch Area-Informationen mit. Alle Netzwerke in den Areas 0 und 2 landen über den nächsten Hop auf 172.18.0.253, also auf BR1. Alle diese Routen wurden über OSPF erlernt und weisen daher ganz links im Eintrag den Buchstaben O auf. Ein weiterer Hinweis darauf, dass es sich um OSPF-Routen handelt, liefert die in eckigen Klammern stehende administrative Distanz von 110. Beispielsweise haben die administrative Distanz und die Metrik für das 172.15.0.0-Netzwerk den Wert [110/3]. Alle mit 192 beginnenden Routen zu den Netzwerken stammen von Area 2. BR4 besitzt kein Interface in Area 2. Es handelt sich also um »Inter-Area«-Routen, was das IA auf der linken Seite bestätigt.

Bewegen wir uns zur Backbone-Area : Die Routingtabelle für BR1 ist in Abbildung 6-17 zu sehen. Beachten Sie, dass die Inter-Area-Routen nun die Routen von Area 1 sind. Ein weiterer Unterschied besteht darin, dass diese Routen nicht an Interfaces, sondern an VLANs gebunden sind. Dieses besondere Gerät ist kein einfacher Router, sondern ein Multilayer-Switch.

Eine der wichtigsten Sachen, die es zu bemerken gilt, ist, dass aufgrund der Positionsänderung verschiedene Ziele über mehrere Pfade erreicht werden können. Das bringt uns zum Thema Schleifen bei OSPF. Gleich der erste Eintrag zeigt, dass zwei Pfade in das Netz-

werk 172.15.0.0 führen. Die Topologie in Abbildung 6-15 stellt tatsächlich noch einen dritten Pfad bereit, doch dieser läuft über Area 2. Der Router hat diesen Pfad zugunsten der kleineren Metrik verworfen. Denken Sie daran, dass OSPF eine dimensionslose Metrik verwendet, die einen Faktor der Netzwerk-Bedingungen darstellt, weshalb die Metrik keine »Einheit« hat. Die beiden Variablen, die diese Routingtabelleneinträge am stärksten beeinflussen, sind die Anzahl der Hops und die Geschwindigkeit des Links. Die vorhin untersuchten Pakete weisen darauf hin, dass keine ToS-Metriken verwendet werden (siehe Abbildung 6-12). Wie wird die Metrik für diesen Eintrag also abgeleitet? Von BR1 aus ist das 172.15.0.0-Netzwerk einen Hop entfernt. Bei Cisco basieren die Kosten dieser Links auf der folgenden Formel: Kosten = 100.000.000 / Link-Bandbreite inbps. Bei einem 100-Mbps-Link (100.000.000-bps) liegen die Kosten also bei 1. Da es zwei Links gibt, über die das Ziel erreichbar ist (Ausgang von BR1 und BR2), liegen die Kosten bei 2.

```
O    172.15.0.0/16 [110/2] via 172.17.0.253, 00:39:11, Vlan17
                   [110/2] via 172.16.0.253, 00:39:11, Vlan16
C    172.17.0.0/16 is directly connected, Vlan17
C    172.16.0.0/16 is directly connected, Vlan16
C    172.18.0.0/16 is directly connected, Vlan18
O    192.168.4.0/24 [110/3] via 192.168.3.253, 23:15:59, Vlan2
     10.0.0.0/16 is subnetted, 2 subnets
O IA    10.2.0.0 [110/2] via 172.18.0.254, 00:27:54, Vlan18
O IA    10.1.0.0 [110/3] via 172.18.0.254, 00:27:55, Vlan18
O    192.168.1.0/24 [110/2] via 192.168.3.253, 23:16:00, Vlan2
O    192.168.2.0/24 [110/3] via 192.168.3.253, 23:16:00, Vlan2
C    192.168.3.0/24 is directly connected, Vlan2
```

▲ Abbildung 6-17
Backbone-Routingtabelle

Bei einer RIP-Topologie wird die 50/50-Lastverteilung erreicht, wenn beide Pfade zum gleichen Ziel die gleiche Anzahl von Hops aufweisen. Die Qualität des Links berücksichtigt RIP dabei nicht. In unseren Fall würde OSPF für Daten, die in Richtung der 172.15.0.0-Netzwerk laufen, das Gleiche machen, da die Metriken identisch sind. Abbildung 6-18 zeigt ein LS-Update-Paket mit erhöhter Metrik. Die Metrik hat sich in diesem Beispiel durch eine Geschwindigkeitskonfiguration am Router-Port geändert. Der dazugehörige Routingtabelleneintrag zeigt nun [110/12] an.

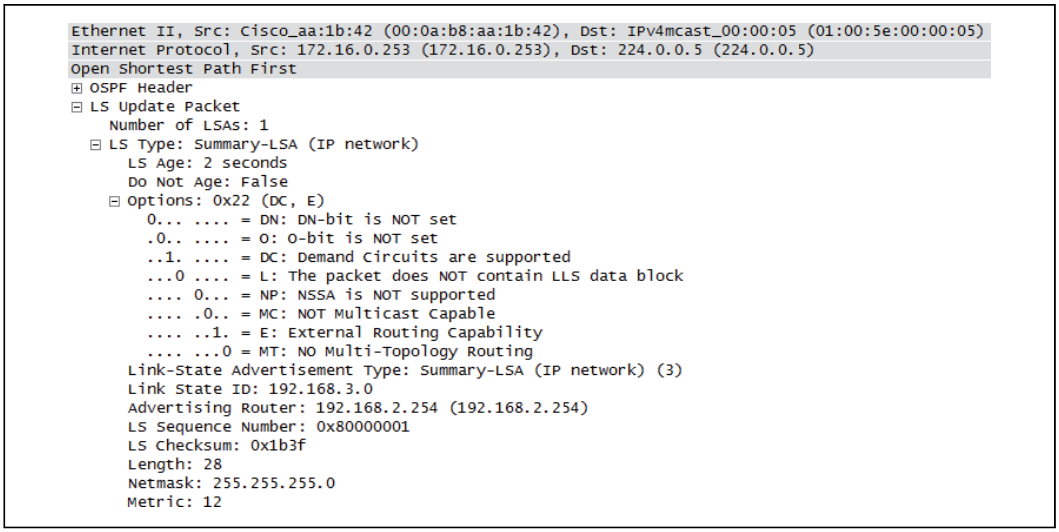


Abbildung 6-18 ▲
LS-Update – Erhöhte Metrik durch Änderung der Linkgeschwindigkeit

Wenn Sie das RIP-Kapitel gelesen haben und wissen, wie RIP funktioniert, können Sie es mit dem Verhalten von OSPF vergleichen. Bei der Betrachtung unterschiedlicher Routingprotokolle wird OSPF als recht aggressiv beschrieben, wenn es darum geht, Pfade zu erkennen und tote Routen zu entfernen. OSPF-Topologien konvergieren und reparieren sich also schneller als RIP-Topologien. Andererseits können OSPF-Updates groß sein und die Hello-Pakete Netzwerk-Ressourcen verbrauchen. Bei stabilen Verhältnissen werden Hello-Pakete bei OSPF häufiger generiert als RIP-Updates. Aufgrund der Verbesserungen nehmen viele Administratoren die Nachteile aber in Kauf.

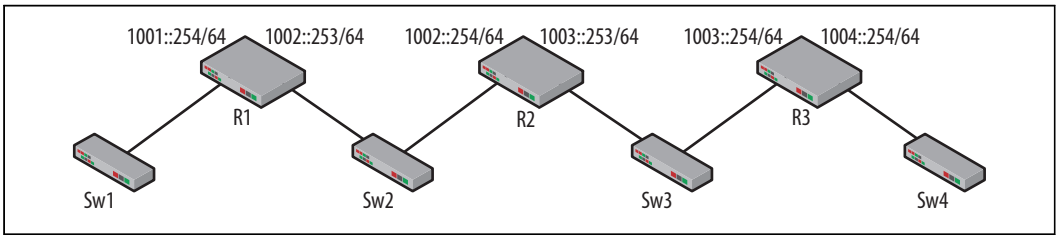
Dass OSPF einen anderen Ansatz verwendet, bedeutet aber nicht, dass es die Lektionen ignoriert, die es von anderen Routingprotokollen gelernt hat. Zum Beispiel besitzt OSPF eine eigene Version von Split Horizon, die Routen nicht in der eigenen Area anbietet. Auch externe Routen werden nicht in allen Areas angeboten. Zusätzlich wird eine Route nicht weiter angeboten, wenn die Metrik den OSPF-Wert für Unendlich überschreitet. Dieser Wert ist ganz unten in Abbildung 6-19 zu sehen. OSPF korrigiert die Topologien mit seinen eigenen getriggerten Updates sehr schnell. Führt ein Link beispielsweise in eine abgeschnittene Area oder gibt es Konfigurationsänderungen, die eine Neuberechnung der LSDB erfordern, sind einige Ziele möglicherweise nicht mehr erreichbar. Wie RIP kann auch OSPF diese Routen als nicht länger erreichbar anzeigen, indem es eine große Metrik verwendet, wie das in Abbildung 6-19 der Fall ist.

```
Ethernet II, Src: Cisco_aa:1b:42 (00:0a:b8:aa:1b:42), Dst: IPv4mcast_00:00:05 (01:00:5e:00:00:05)
Internet Protocol, Src: 172.16.0.253 (172.16.0.253), Dst: 224.0.0.5 (224.0.0.5)
Open Shortest Path First
⊞ OSPF Header
⊟ LS Update Packet
    Number of LSAs: 1
  ⊟ LS Type: Summary-LSA (IP network)
      LS Age: 3600 seconds
      Do Not Age: False
    ⊟ Options: 0x22 (DC, E)
        0... .... = DN: DN-bit is NOT set
        .0.. .... = O: O-bit is NOT set
        ..1. .... = DC: Demand Circuits are supported
        ...0 .... = L: The packet does NOT contain LLS data block
        .... 0... = NP: NSSA is NOT supported
        .... .0.. = MC: NOT Multicast Capable
        .... ..1. = E: External Routing Capability
        .... ...0 = MT: NO Multi-Topology Routing
      Link-State Advertisement Type: Summary-LSA (IP network) (3)
      Link State ID: 192.168.1.0
      Advertising Router: 192.168.2.254 (192.168.2.254)
      LS Sequence Number: 0x80000002
      LS Checksum: 0xb6b0
      Length: 28
      Netmask: 255.255.255.0
      Metric: 16777215
```

▲ **Abbildung 6-19**
Unendlich-Metrik bei OSPF

OSPF und IPv6

OSPF für IPv6 ist der IPv4-Version ähnlich. Das Verhalten und der Paket-Traffic bleiben gleich. Die Datenbankbeschreibung ist etwas komplexer, und die Pakete wurden um zusätzliche Optionen und LSA-Typen erweitert. Die Topologie in Abbildung 6-20 zeigt die IPv6-Topologie für diesen Abschnitt.



▲ **Abbildung 6-20**
IPv6-Topologie

Wie im letzten Kapitel gezeigt wurde, sind bei der Konfiguration des Routingprotokolls für IPv6 viele Befehle an das Interface gebunden. Eine grundlegende IPv6-OSPF-Konfiguration sieht wie folgt aus:

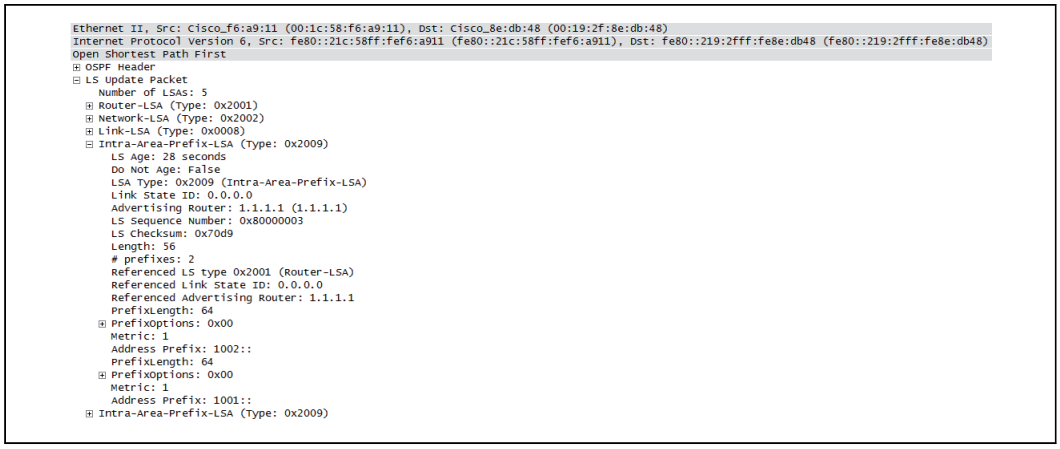
```
ipv6 unicast-routing
interface FastEthernet0/0
    ipv6 address 1001::254/64
    ipv6 ospf 1 area 51
interface FastEthernet0/1
    ipv6 address 1002::253/64
    ipv6 ospf 1 area 51
ipv6 router ospf 1
    router-id 1.1.1.1
```

Die Router-ID wird zur Identifizierung des Routers verwendet, für den Fall, dass keine IPv4-Adressen auf dem Router präsent sind. Sobald alle Router konfiguriert wurden, werden auch die Routingtabellen um die per OSPF gelernten Routen aktualisiert (siehe Abbildung 6-21). Um die Komplexität dieses Beispiels zu reduzieren, wurde nur eine Area konfiguriert.

Abbildung 6-21 ▶
IPv6-OSPF-Routingtabelle

Wie bei den anderen IPv6-Beispielen in diesem Buch auch, sind die direkt verbundenen Routen und die Link-Routen vom Routingprotokoll unabhängig. Die OSPF-Routen wurden installiert und weisen nahezu die gleichen Informationen auf wie ihre IPv4-Gegenstücke: Netzwerkadresse, Netzwerkmaske, administrative Distanz und Metrik.

Alle IPv6-Pakettypen durchzugehen würde den Rahmen dieses Kapitels sprengen, aber es lohnt sich, einen Blick auf ein Beispiel-Paket zu werfen. Abbildung 6-22 zeigt ein LS-Update, das zwischen R1 und R2 abgefangen wurde. Die IPv6-LS-Updates können recht umfangreich sein, weshalb ein Großteil der Header nicht zu sehen ist. Die dargestellten Felder zeigen das Netzwerk-Präfix, die Maskenlänge und die Verwendung der Router-ID. Wie bereits erwähnt, wurden zusätzliche Optionen (wie etwa ein Flag für IPv6) und LSA-Typen hinzugefügt.



▲ **Abbildung 6-22**
IPv6-OSPF-LS-Update

In diesem Beispiel ist einer der neuen LSA-Typen (Intra-Area-Präfix-LSA) zu sehen. RFC 5340 sagt dazu (frei übersetzt) Folgendes:

> Dieses LSA enthält alle IPv6-Präfix-Informationen, die bei IPv4 in Router-LSAs und Netzwerk-LSAs enthalten sind.

Das bedeutet auch, dass die Netzwerk- und Router-LSAs bei IPv6 geändert wurden und Adressinformationen nicht mehr übertragen.

Lektüre

RFC 1584: »Multicast Extensions to OSPF«
RFC 1587: »The OSPF NSSA Option«
RFC 1793: »Extending OSPF to Support Demand Circuits«
RFC 2328: »OSPF version 2 (obsoletes RFCs 2178, 1583, 1247)«
RFC 2370: »The OSPF Opaque LSA Option«
RFC 2547: »Using an LSA Options Bit to Prevent Looping in BGP/MPLS IP VPNs«
RFC 3883: »Detecting Inactive Neighbors over OSPF Demand Circuits (DC)«
RFC 4811: »OSPF Out-of-Band Link State Database (LSDB) Resynchronization«
RFC 4915: »Multi-Topology (MT) Routing in OSPF«
RFC 5340: »OSPF for IPv6 (obsoletes RFC 2740)«
RFC 5613: »OSPF Link Local Signaling (obsoletes RFC 4813)«

Zusammenfassung

Als Link-State-Protokoll gibt es beim Betrieb von OSPF viele wichtige Änderungen gegenüber RIP. Hello-Nachrichten, Link-State-Updates, Datenbankbeschreibungen, zusätzliche Metriken und die schnelle Verbreitung von Routinginformationen dienen alle zur Steigerung der Performance. OSPF wird hierarchisch betrieben, d.h., die Router verstehen die Topologie besser, wenn sie zu Area-Border- oder Backbone-Routern aufsteigen. Router innerhalb der Areas wissen nur sehr wenig vom gesamten autonomen System.

Es erfordert recht viel Erfahrung, um ein OSPF-Experte zu werden. Zwar sind das Einrichten und der Betrieb recht einfach, allerdings ist OSPF deutlich komplexer als Distanzvektor-Protokolle. Dennoch sollten Sie als Leser dank der Paket-Captures, der Definitionen und unserer Betrachtungen des Betriebs gut gerüstet sein, um OSPF-Topologien konfigurieren und Fehler aufspüren zu können.

Fragen

1. Wählen Sie die Begriffe aus, die OSPF am besten beschreiben.
   a. Link State, flach
   b. Link State, hierarchisch
   c. Distanzvektor, flach
   d. Distanzvektor, hierarchisch
2. Wie lautet die Ziel-IP-Adresse für OSPF?
3. Bei OSPF-Topologien glaubt jeder Router anfänglich, er sei die Wurzel.
   a. WAHR
   b. FALSCH
4. OSPF-Area Null ist als externe Route bekannt.
   a. WAHR
   b. FALSCH
5. Bei OSPF wissen die Router innerhalb einer Area nichts über die AS-Topologie.
   a. WAHR
   b. FALSCH
6. Welchen Wert hat der OSPF-Hello-Timer?

7. Die Netzwerk-LSA wird von allen Routern verwendet.
   a. WAHR
   b. FALSCH
8. Welche OSPF-Nachricht ist vom Typ 4?
   a. DB-Beschreibung
   b. LS-Update
   c. LS-ACK
   d. LS-Request
9. Welche administrative Distanz hat OSPF?
10. Aus welchen beiden Werten setzt sich die Standard-Metrik von OSPF zusammen?

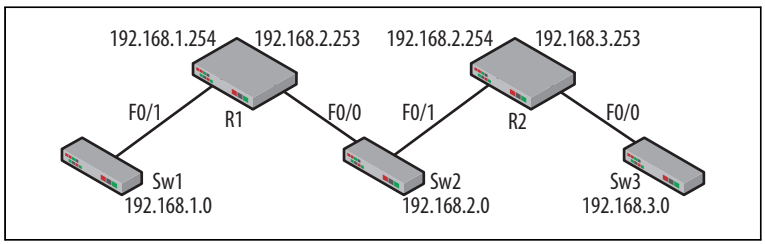
Antworten

1. Link-State, hierarchisch
2. 224.0.0.5
3. WAHR
4. FALSCH
5. WAHR
6. 10 Sekunden
7. FALSCH
8. LS-Update
9. 110
10. Geschwindigkeit des Links und Anzahl der Hops

Laborübungen

Übung 1: Aufbau der Topologie aus Abbildung 6-23

Material: 2 Router, 2 Computer, optional Switches (oder VLANs) für jedes Netzwerk



◀ **Abbildung 6-23** Topologie für Experiment 1

1. Verkabeln Sie die Topologie, und konfigurieren Sie die IP-Adressen an den Router-Interfaces.
2. Verbinden Sie jeweils einen Computer mit den Netzwerken 192.168.1.0 und 192.168.2.0.
3. Konfigurieren Sie manuell die IP-Adressen und Gateways für die Computer.
4. Spielt es für den Computer im 192.168.2.0-Netzwerk eine Rolle, welches Standard-Gateway verwendet wird? Warum? Wie sieht das aus, wenn OSPF läuft?
5. Untersuchen Sie die Routingtabellen auf den Routern. Was sehen Sie? Praktischer Cisco-Befehl: `show ip route`

Übung 2: OSPF auf den Routern aktivieren

Material: Topologie aus Übung 1, Wireshark

1. Konfigurieren Sie OSPF auf jedem Router.
2. Praktische Cisco-Befehle: `router ospf process-id`, `network _____ area _____`
3. Halten Sie Traffic auf beiden Rechnern fest, und schauen Sie sich an, wie die OSPF-Pakete zu fließen beginnen. Sobald die Konfiguration abgeschlossen ist, untersuchen Sie erneut die Routingtabellen der Router.
4. Was hat sich an den Routingtabellen geändert. Welche Werte stehen in den eckigen Klammern? Warum?

Übung 3: Tracing des Paketflusses

Material: Topologie aus Übung 1, Wireshark

1. Koppeln Sie die Router-Interfaces von der Topologie ab.
2. Stellen Sie sicher, dass der Wireshark-Capture im 192.168.2.0-Netzwerk läuft.
3. Untersuchen Sie die festgehaltenen Pakete, und verfolgen Sie den Paketaustausch. Können Sie die Reihenfolge der Pakete und die Gründe für jedes Paket identifizieren?
4. Wählen Sie ein Link-State-Update mit Routinginformationen. Können Sie die in diesem Kapitel beschriebenen Felder identifizieren?

Übung 4: Netzwerk-Bedingungen ändern

Material: Topologie aus Übung 1, Wireshark

1. Sagen Sie vorher, was passiert, wenn sich die IP-Adresse für F0/0 an R2 ändert. Welche Pakete könnten erzeugt und welche Änderungen an den Routingtabellen vorgenommen werden?
2. Bei laufendem Wireshark ändern Sie die IP-Adresse in 192.168.4.253. Führen Sie auch die notwendigen Änderungen an der OSPF-Konfiguration durch.
3. Welcher Traffic wird daraufhin erzeugt? Welche Änderungen haben die Routingtabellen erfahren? Wie nah dran waren Ihre Vorhersagen?
4. Stellen Sie für die nächste Übung die Ur-Topologie wieder her.

Übung 5: Eine Schleife

Material: Topologie aus Übung 1, Wireshark, ein Switch zwischen R1 und R2

1. Ändern Sie die IP-Adresse von F0/0 an R2 in 192.168.1.253.
2. Verbinden Sie das Interface mit Switch 1, wodurch eine Schleife aufgebaut wird.
3. Welche Änderungen werden an den Routingtabellen vorgenommen?
4. Mit der aufgebauten Schleife und laufendem Wireshark pingen Sie die Adressen des 192.168.1.0-Netzwerks von R2 an. Über welche Links läuft der Traffic? Wie handhabt OSPF die Lastverteilung?

Über den Autor

Bruce Hartpence arbeitet am Network, Security, and System Administration Department (NSSA) des Rochester Institute of Technology (RIT), New York. Er verbringt gleichermaßen viel Zeit mit dem Unterrichten wie mit der Durchführung von Projekten und mit dem Schreiben von Fachpublikationen.

Über den Übersetzer

Peter Klicman ist unabhängiger Sachverständiger für DV-Systeme sowie Internet-Provider und freier Unternehmensberater. Seine Arbeit für den O'Reilly Verlag brachte ihn zur technischen Dokumentation. Neben Buchübersetzungen führt er Dokumentations- und Entwicklungsprojekte durch.

Kolophon

Auf dem Cover von »Praxiskurs Routing und Switching« ist ein Grünes Heupferd (Tettigonia viridissima) abgebildet. Diese zu den Laubheuschrecken gehörende Art kommt in ganz Mitteleuropa und Asien bis zum Pazifik vor. Die Färbung ist grün, mit einem bräunlichen Streifen auf dem Rücken. Da das Grüne Heupferd zu der Ordnung der Langfühlerschrecken gehört, stehen von seinem Kopf zwei lange Fühler nach vorne ab. Mit drei bis vier Zentimetern Rumpflänge sind diese Insekten auffallend groß. Aus dem Hinterleib der Weibchen ragen noch einmal zwei bis drei Zentimeter lange Legeröhren heraus. Die Flügel, mit denen sie im Gegensatz zu anderen Heuschrecken auch recht gut fliegen können, reichen beim ausgewachsenen Tier über den Körper hinaus. Mit ihren großen Kauwerkzeugen ernähren sie sich überwiegend räuberisch von Insekten und Larven, verschmähen auch eigene verletzte Artgenossen nicht, fressen aber auch Gräser und Kräuter.

Schon aus hundert Meter Entfernung kann man den lauten Gesang der Männchen ausmachen. Dieser entsteht durch das Aneinanderreiben der Flügel. Dort befindet sich das Stridulationsorgan – eine Membran, die durch Schrillleisten und Schrillzähne zum Schwingen gebracht wird. Um den Gesang der anderen Tiere wahrzunehmen, verwenden sie ihr Gehörorgan in zwei kleinen Gruben auf den Vorderbeinen, das so genannte Tympanalorgan.

Je nach Trockenheit der Sommer benötigen Heupferde ein bis fünf Jahre, um sich nach der Eiablage zum adulten Tier zu entwickeln. In den ersten Jahren liegen die Eier bis zum Schlüpfen der Larven in der Erde. Innerhalb von drei Monaten häuten sich die Larven siebenmal und verbringen den Spätsommer und Herbst als ausgewachsene Tiere. Ihr Gesang ist an warmen Oktobertagen noch bis in die späten Abendstunden zu hören.

Das Coverlayout dieses Buchs hat Michael Oreal gestaltet. Als Textschrift verwenden wir die Linotype Birka, die Überschriftenschrift ist die Adobe Myriad Condensed, und die Nichtproportionalschrift für Codes ist LucasFont's TheSansMono Condensed.

Index

P

R

S

T

U

V

W

Y

Z