

6 Routing

6.1 IP-Protokoll

6.2 Ziele der Wegewahl

6.3 Ein Graphenmodell

6.4 Ein Routermodell

6.5 Kürzeste Wege in Graphen

- a) Algorithmus von Dijkstra
- b) Algorithmus von Bellman und Ford

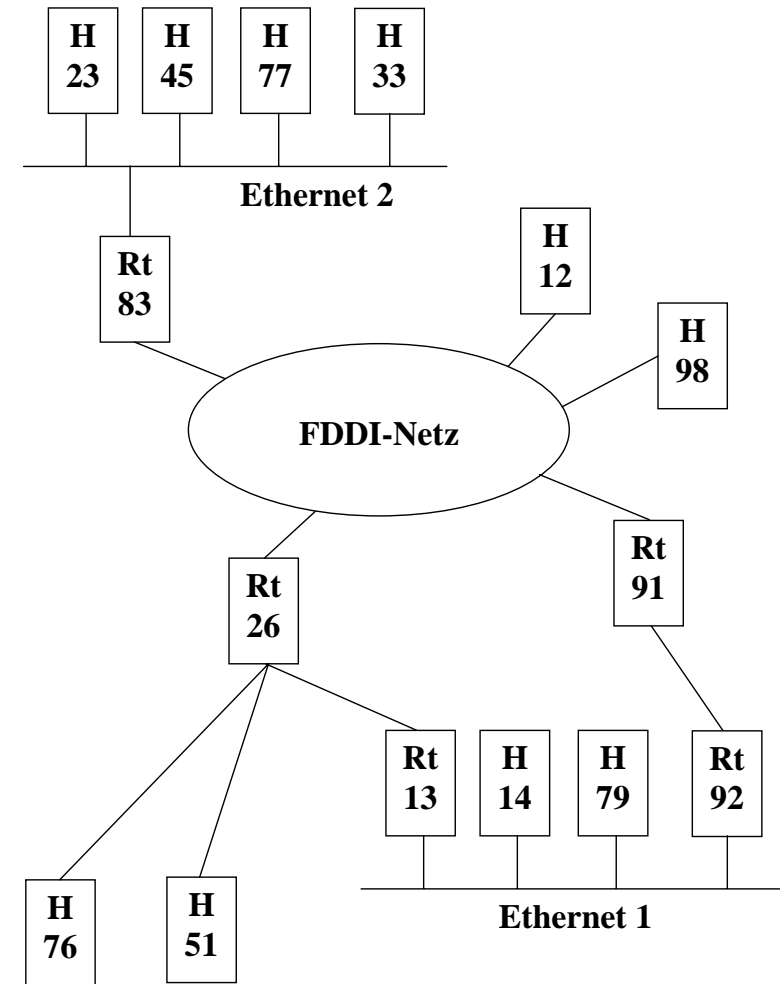
6.6 Adaptives Routing

6.7 Multicast-Routing

6.8 Hierarchisches Routing

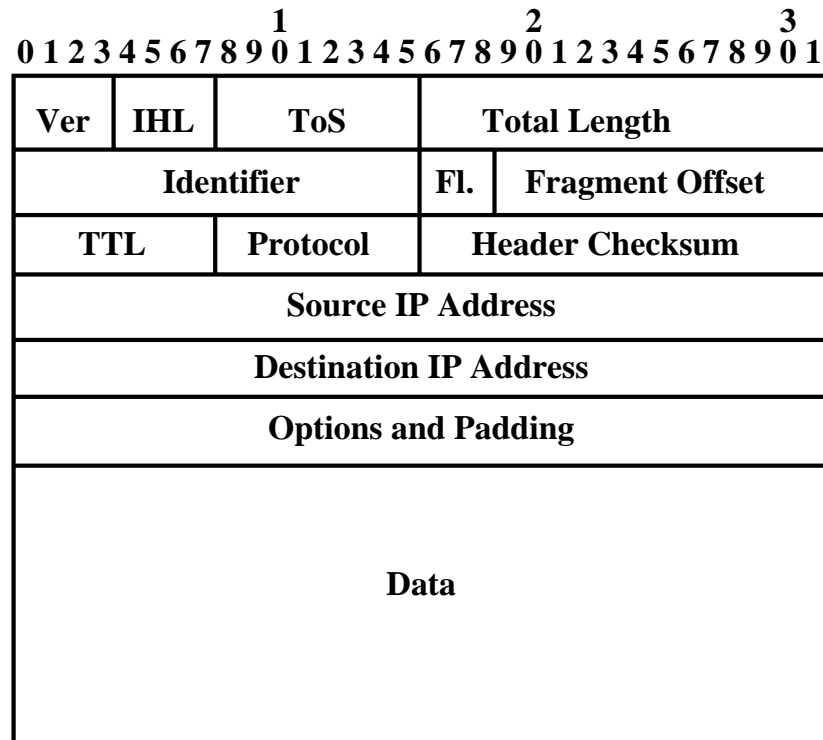
6.9 Abschließende Bemerkungen

Bild eines Netzes von Netzen:



Legende: H = Host, Rt = Router

Format eines Internet Protokoll Pakets:



Erläuterungen zu den einzelnen Felder:

Ver = Version (4 Bit) = Versionsnummer des IP-Protokolls für dieses Datagramm, die Versionsnummer stellt sicher, daß alle an einem Datentransfer beteiligten Instanzen das Paketformat gleich interpretieren.

IHL = Internet Header Length (4 Bit) = Länge des Kopfes in 32-Bit-Einheiten; aufgrund des Optionsfeldes kann ein Kopf variabel lang sein, Minimallänge ist 5.

ToS = Type of Service (8 Bit): Spezifiziert Verlässlichkeits-, Vorrang- und Durchsatzstufen; Aufbau des Feldes: PPPDTRUU, wobei PPP die Wichtigkeit des Datagramms in Stufen von 0 bis 7 angibt, 0 steht für normale Datagramme, 7 für Datagramme zur Verwaltung des Netzes; D, T, und R beschreiben den gewünschten Transportservice, D steht für Weg geringer Transportdauer, T für Weg hohen Durchsatzes und R für Weg hoher Verlässlichkeit; U bedeutet ungenutzt.

Total Length (16 Bit): Länge des gesamten Datagramms in Oktetten, die Länge des Datenbereiches berechnet man als Differenz aus Total Length und $4 * IHL$.

Identifier (16 Bit): Das Tripel "Source IP Address", "Destination IP Address" und "Identifier" kennzeichnet eindeutig ein Datagramm während seiner Lebenszeit, dies wird bei notwendigen Fragmentierungen eines Datagramms genutzt.

Fl. = Flags (3 Bit): Aufbau des Feldes: UMD, wobei U für ungenutzt, M für "More-Fragments" und D für "Don't-Fragment" steht.

Fragment Offset (13 Bit): Position des Dateninhaltes in 64-Bit-Einheiten des aktuellen Fragments bezüglich des Originalpakets.

TTL = Time to Live (8 Bit): Eine Ganzzahl, die in jedem Router erniedrigt wird. Bei Erreichen von 0 wird das Datagramm vernichtet. Ursprünglich beschrieb TTL die Lebensdauer in Sekunden.

Protocol (8 Bit): Kennzeichen des Protokolls, das für die Erstellung des Datenfeldes genutzt wurde.

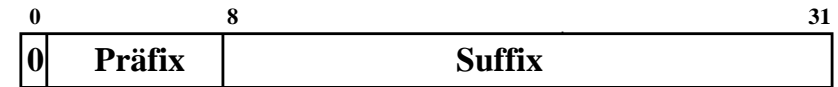
Header Checksum (16 Bit): Es wird die Summe S aller 16-Bit-Einheiten des Kopfes mit Ausnahme des Feldes Header Checksum in Einerkomplement-Arithmetik berechnet, gespeichert wird das Einerkomplement von S.

Options and Padding: Dieses Feld wird u. a. genutzt, um Sicherheitsanforderungen an Zwischennetze zu benennen, um Transportwege für ein Datagramm vorzuschreiben, um Transportwege und/oder Transportdauern eines Datagramms aufzuzeichnen. Es ist zu beachten, daß die Oktettzahl dieses Feldes ein ganzzahliges Vielfaches von 4 ist.

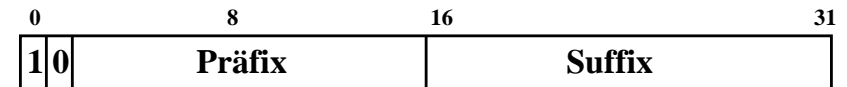
Bemerkung: In RFC 1349 wurde das ToS-Feld um ein M-Bit für Routenwahl gemäß minimaler Kosten ergänzt. In RFC 2474 wurde das ToS-Feld durch ein DS-Feld ersetzt. DS steht für Differentiated Services. Bei der Neudefinition wurde darauf geachtet, daß die neue Definition zur ursprünglichen Definition weitgehend kompatibel ist.

IP-Adreßklassen:

"Class A"-Adresse:



"Class B"-Adresse:



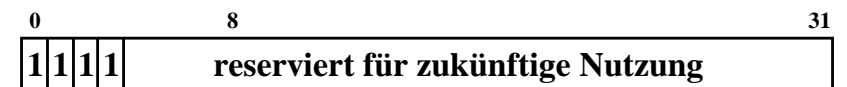
"Class C"-Adresse:



"Class D"-Adresse:



"Class E"-Adresse:



Bemerkung: Das Präfix kann man als Netzadresse und das Suffix als Hostadresse deuten, in OSI-Terminologie wird eine IP-Adresse als NSAP (Network Service Access Point) bezeichnet. Das Klassensystem für Adressen wurde ersetzt durch ein klassenloses System.

Schreibweise der Adressen mittels Punkten:

Die vier Oktette einer IP-Adresse werden zwecks besserer Lesbarkeit durch Punkte getrennt, so steht z. B. 192.5.48.3 für die 32-Bitzahl 1100000000000101001100000000011, in diesem Fall handelt es sich um eine Adresse der Klasse C.

Unterteilung des Adreßraumes:

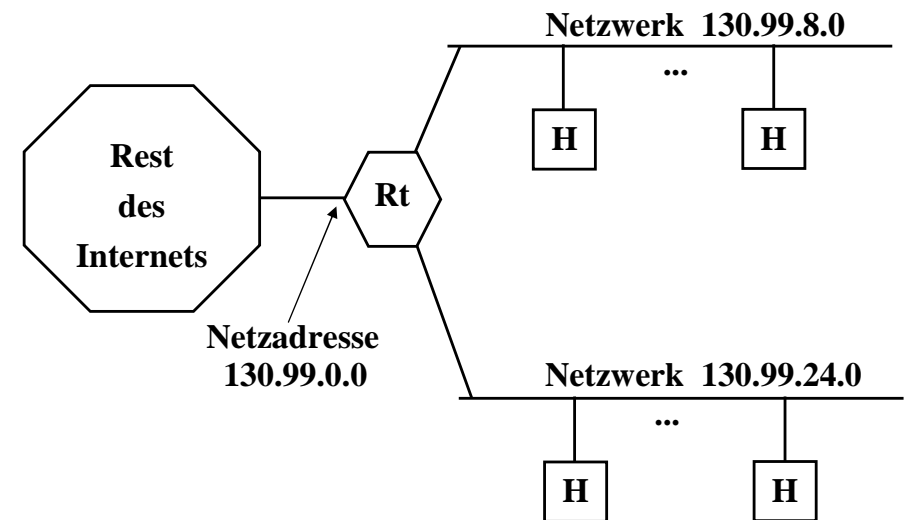
Adreß-klasse	Präfix-länge	Maximale Netzzahl	Suffix-länge	Maximale Hostzahl
A	7 Bit	128	24 Bit	16777216
B	14 Bit	16384	16 Bit	65536
C	21 Bit	2097152	8 Bit	256

IP-Adressen mit Sonderbedeutung:

Präfix	Suffix	Adreßbedeutung
0...0	0...0	"dieser Rechner während BV"
0...0	hh	Host hh in diesem Netz
nn	0...0	Netzwerk nn
nn	1...1	gerichteter Rundspruch in nn
1...1	1...1	lokaler Rundspruch
127	≠ 0...0	Echoadresse für Testzwecke

Bemerkung: Die obigen Spezialadressen dürfen nur für den angegebenen Zweck genutzt werden; sie sind niemals eigenständige Hostadressen.

Bildung von Unternetzen:



Der Router Rt entscheidet durch Betrachtung des dritten Oktetts, welchem Netz ein Datagramm übergeben wird, dort wird es von einem lokalen Router zu seinem Bestimmungsort weitergeleitet.

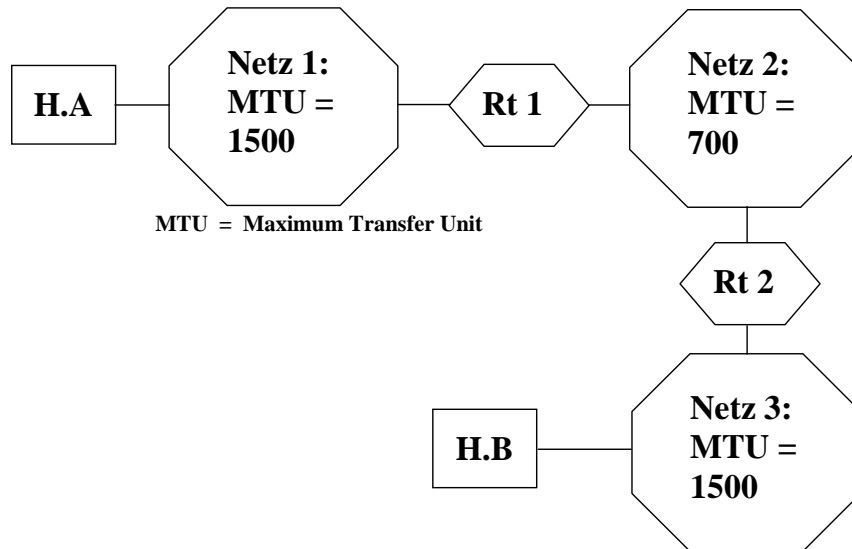
Unternetze beschreibt man durch Bitmasken, wobei empfohlen wird, den Netzteil einer Adresse durch eine zusammenhängende Folge von Einerbit zu kennzeichnen. Ein Beispiel:

11111111.11111111.11111000.000000

Netzzadresteil Hostadresteil

Analog zur Unterteilung einer Klasse B Adresse läßt sich ein Bereich von Klasse C Adressen zu einer Supernetzadresse zusammenfassen. So kann man die obige Maske auch zur Beschreibung des Adreßbereichs 234.170.168.0 bis 234.170.175.255 verwenden.

Beispiel zur Fragmentierung:



Ursprüngliches Datagramm von H.A nach H.B:

Kopf 1	1400 Oktette
--------	--------------

Vom Router Rt.1 erzeugte Fragmente:

Kopf 2	700-20 Oktette	Fragment Offset = 0
Kopf 3	680 Oktette	Fragment Offset = $680/8 = 85$
Kopf 4	40 Oktette	Fragment Offset = $1360/8 = 170$

Bemerkungen:

- (i) Die Köpfe 1, 2, 3 und 4 stimmen weitgehend überein.
- (ii) Fragmentierungen können mehrfach erfolgen.
- (iii) Erst im Ziel wird das Ursprungsdatagramm wiederhergestellt.

Internet Control Message Protocol (ICMP):

ICMP ist das Bruderprotokoll zu IP. Während in IP das normale Übersenden von Datagrammen geregelt wird, ist ICMP für Sondersituationen, insbesondere Fehlersituationen, zuständig.

Teilliste von ICMP-Nachrichten:

Echo Reply,
Destination Unreachable,
Source Quench,
Redirect,
Echo,
Router Advertisement,
Router Solicitation,
Time Exceeded,
Parameter Problem,
Timestamp,
Timestamp Reply,
Information Request,
Information Reply,
Address Mask Request,
Address Mask Reply,
Traceroute,
Datagram Conversion Error,
Domain Name Request,
Domain Name Reply.

Ziel der Wegewahl: Gute Nutzung des Netzes

Maße: Verzögerungszeit,
Durchsatz,
Sicherheit, ...

Bemerkung: Die Wegewahl ist eng verknüpft mit Fragen der Fluß- und Staukontrolle.

Beispiel:

Man habe zwei Klassen von Nachrichten:

- (a) kurze Nachrichten (< 500 Bit)
- (b) lange Nachrichten ($> 10^7$ Bit)

z. B. Weg für lange Nachrichten:



Der Wunsch nach fairer und effizienter Nutzung des Übertragungsweges führt zu zwei Rückmeldungen: Quittung und Freimeldung.

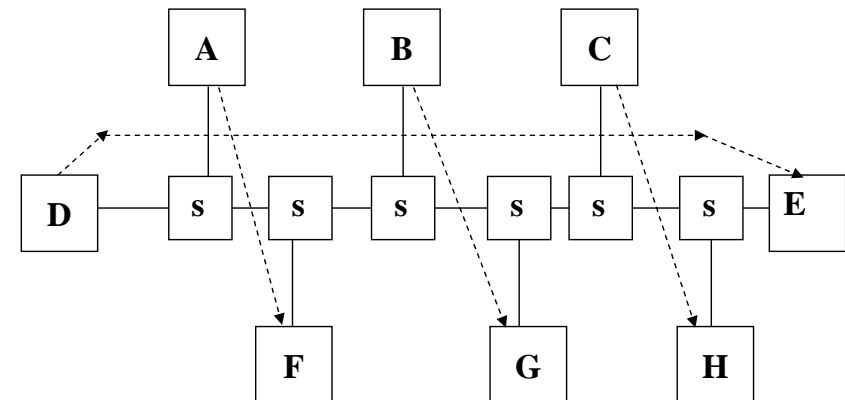
Kriterien für Routing-Algorithmen:

- Korrektheit
- Robustheit
- Stabilität
- Fairness
- Einfachheit
- Optimalität

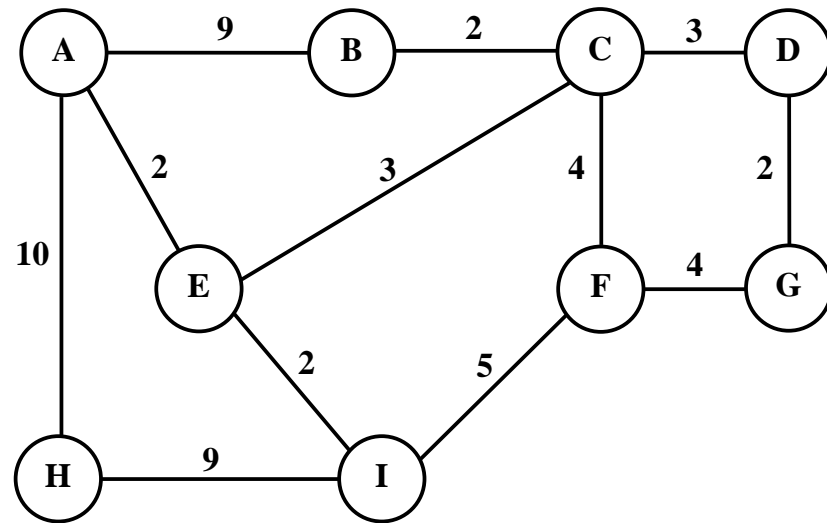
Konflikt zwischen Fairness und Optimalität:

Es stehen Übertragungen an von A nach F, von B nach G, von C nach H, die jeweils die gesamte Kapazität des entsprechenden Leitungsabschnittes zwischen D nach E in Anspruch nehmen, und von D nach E.

Frage: Soll Übertragung von D nach E verschoben werden?



Ein Transportnetz als kantenbewerteter Graph:

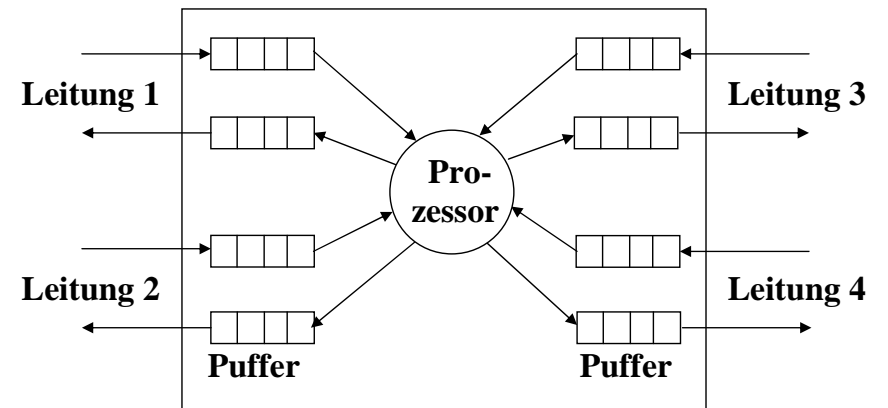


Bemerkungen:

Der obige Graph ist ungerichtet. Zwischen je zwei Knoten existieren mindestens zwei unabhängige Wege. Die Zahlen an den Kanten können verschieden interpretiert werden, z. B. als geographische Entfernungen, Leitungskapazitäten, Transportkosten je Paket, momentanes Verkehrsaufkommen zwischen den einzelnen Knoten, Störanfälligkeit von Leitungen.

Sieht man in den Kantenbewertungen Kosten, dann läßt sich ein kostengünstigster Weg zwischen je zwei Knoten berechnen.

Modell eines Routers:



Informationsquelle für Routing:

- lokal
- lokale Umgebung des Routers
- Gesamtnetz

Verfahrensarten:

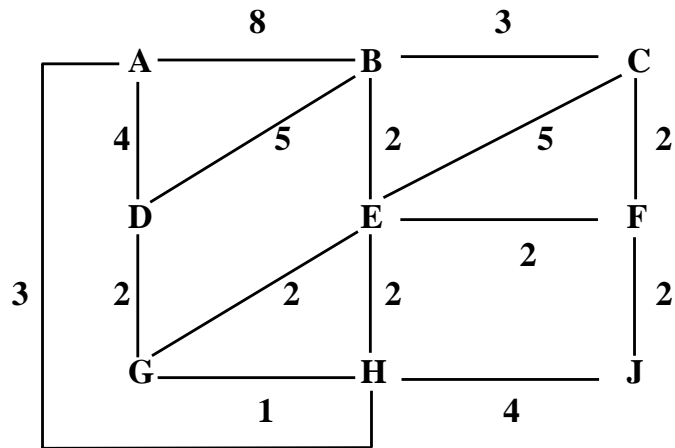
- starr
- adaptiv

Beispiele für Routingverfahren:

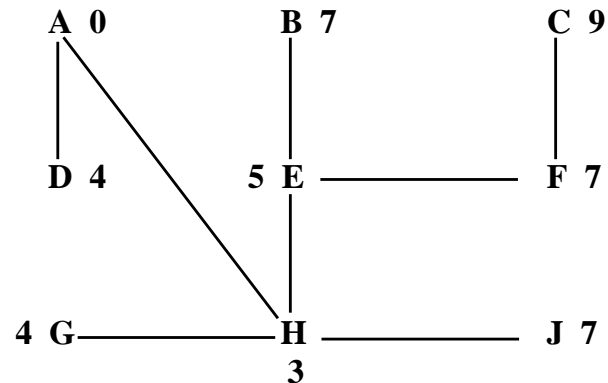
- Fluten
- idealer Beobachter
- "random routing"
- "hot potato"-Verfahren
- Barans heuristisches "backward learning"
- kürzeste Wege

Beispiel zu Dijkstras Algorithmus:

Bewerteter Graph:



Wegebaum für A:



Berechnung eines Wegebaums für Quelle A mittels Ausschöpfung:

Menge	Neuer Knoten	Weglängen							
		B	C	D	E	F	G	H	J
{A}	-	8	∞	4	∞	∞	∞	3	∞
		<i>wähle Minimalknoten, d.h. H</i>							
{A,H}	H	8	∞	4	5	∞	4	3	7
		<i>wähle Minimalknoten, d.h. D</i>							
{A,D,H}	D	8	∞	4	5	∞	4	3	7
		<i>wähle Minimalknoten, d.h. G</i>							
{A,D,G,H}	G	8	∞	4	5	∞	4	3	7
		<i>wähle Minimalknoten, d.h. E</i>							
{A,D,E,G,H}	E	7	10	4	5	7	4	3	7
		<i>wähle Minimalknoten, d.h. B</i>							
{A,B,D,E,G,H}	B	7	10	4	5	7	4	3	7
		<i>wähle Minimalknoten, d.h. F</i>							
{A,B,D,E,F,G,H}	F	7	9	4	5	7	4	3	7
		<i>wähle Minimalknoten, d.h. J</i>							
{A,B,D,E,F,G,H,J}	J	7	9	4	5	7	4	3	7
		<i>wähle Minimalknoten, d.h. C</i>							
{A,B,C,D,E,F,G,H,J}	C	7	9	4	5	7	4	3	7

Fertig!

Wegetabelle für Knoten A:

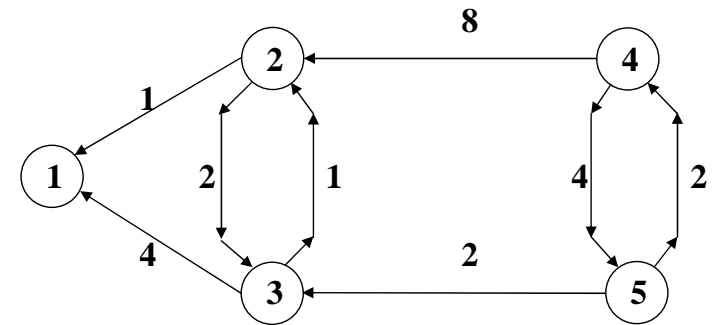
Ziel	Wert	Leitung
A	0	-
B	7	H
C	9	H
D	4	D
E	5	H
F	7	H
G	4	H
H	3	H
J	7	H

Analog berechnet man die übrigen Wegetabellen, z. B. für Knoten H.

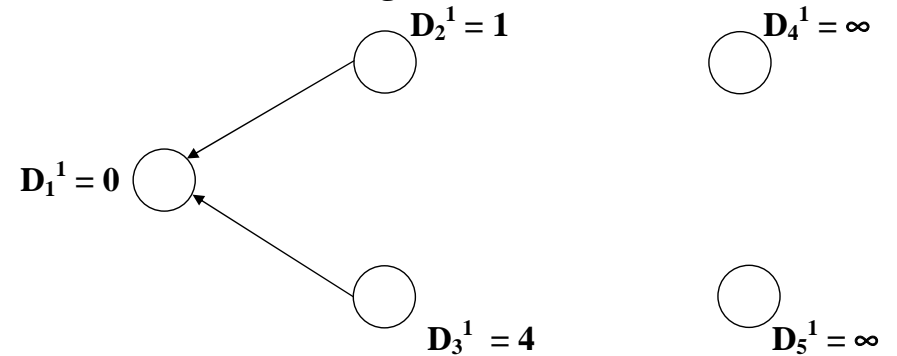
Wegetabelle für Knoten H:

Ziel	Wert	Leitung
A	3	A
B	4	E
C	6	E
D	3	G
E	2	E
F	4	E
G	1	G
H	0	-
J	4	J

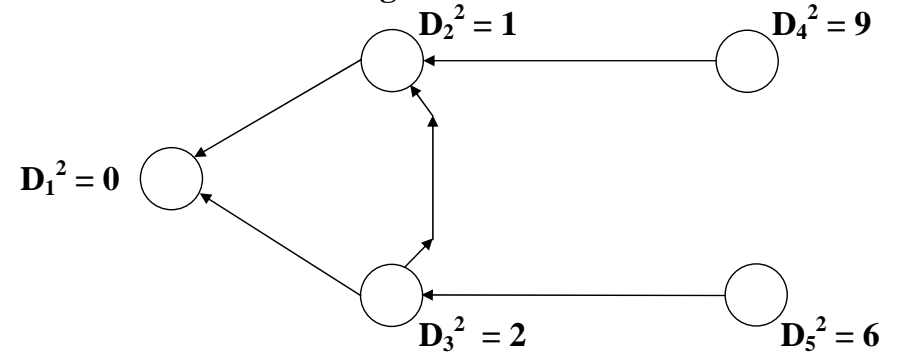
Beispiel zum Bellman-Ford Algorithmus:



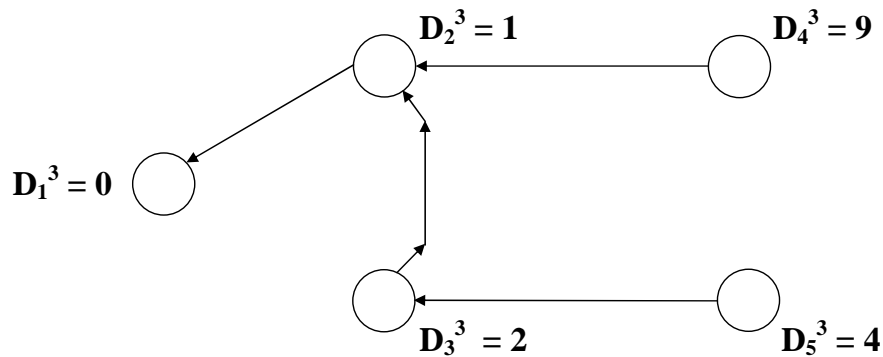
Schritt 1: Höchstens Wege der Kantenzahl 1:



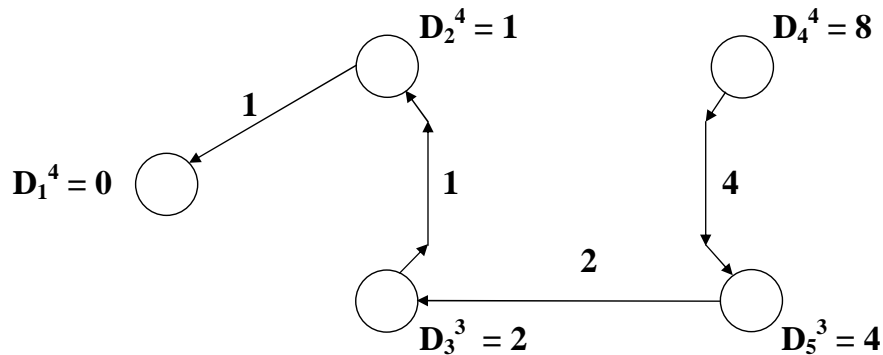
Schritt 2: Höchstens Wege der Kantenzahl 2:



Schritt 3: Höchstens Wege der Kantenzahl 3:



Endschritt: Höchstens Wege der Kantenzahl 4:



Bemerkung: Der Algorithmus von Bellman und Ford konvergiert unter recht allgemeinen Voraussetzungen; er ist als verteilter Algorithmus einsetzbar; eine Randbedingung besteht darin, daß das Netz keine negativen Zyklen enthält.

Barans "backwards learning":

Voraussetzung:

Übertragungszeit von Rechner I nach Rechner K =
Übertragungszeit von Rechner K nach Rechner I

Jeder Rechner unterhält eine Tabelle von Schätzwerten der Transitzeiten zu jedem Ziel über jeden Port.

		Port						
		1	2	3	4	.	.	.
Ziel:	A	12,3	20,6	35,8	14,7
	B	66,9	80,3	56,2	67,1
	C	17,4	90,2	98,8	80,7
	D
						

Jedes von der Quelle Q über den Port P eintreffende Paket führt zu einer Aktualisierung des Schätzwertes $S(Q, P)$, sei v die gemessene Verzögerungszeit des betrachteten Paketes, dann

$$S_{\text{neu}}(Q, P) = a * S_{\text{alt}}(Q, P) + b * v$$

mit $a + b = 1$.

Bemerkung: Eine geschickte Wahl von a mindert Schwingungseffekte.

Beispiel für adaptives Routing:

Austausch von Schätzungen über Transitzeiten

Alte Routing-Tabelle für Knoten 1:

Zielknoten	Zeit	Ausgang
1	0	-
2	2	2
3	5	3
4	1	4
5	6	3
6	8	3

Schätzwerte der Nachbarn von Knoten 1:

von 2		von 3		von 4	
Ziel	Zeit	Ziel	Zeit	Ziel	Zeit
1	2	1	3	1	1
2	0	2	3	2	2
3	3	3	0	3	2
4	2	4	2	4	0
5	3	5	1	5	1
6	5	6	3	6	3

Neue Zeitwerte für den Transit zu den Nachbarn:

- Nachbar 2: 2
- Nachbar 3: 3
- Nachbar 4: 1

Neue Routing-Tabelle für Knoten 1:

Zielknoten	Zeit	Ausgang	
1	0	-	
2	2	2	
3	3	4	← Nicht der direkte Weg nach Knoten 3 !
4	1	4	
5	2	4	
6	4	4	

Berechnungsverfahren:

Wähle als neuen Schätzwert in der Routing-Tabelle des Routers a für den Zielknoten z den Wert M und den Ausgang i, für den $M(j) = \text{Transitzeit}(a, j) + \text{Transitzeit}(j, z)$ minimal bezüglich der Nachbarknoten j von a ist.

Beispiel: $\text{Transitzeit}(1, 2) + \text{Transitzeit}(2, 6) = 2 + 5 = 7,$
 $\text{Transitzeit}(1, 3) + \text{Transitzeit}(3, 6) = 3 + 3 = 6,$
 $\text{Transitzeit}(1, 4) + \text{Transitzeit}(4, 6) = 1 + 3 = 4,$
 damit $M(4) = 4.$

Beispiel zum adaptiven verteilten Routing:

- "Gute" Nachrichten verbreiten sich schnell:
Router A wird wieder aktiv.

Abstand der Router zu A

	A	B	C	D	E
0:	-	∞	∞	∞	∞
1:	0	1	∞	∞	∞
2:	0	1	2	∞	∞
3:	0	1	2	3	∞
4:	0	1	2	3	4

- "Schlechte" Nachrichten werden zögernd akzeptiert:
Router A geht vom Netz.

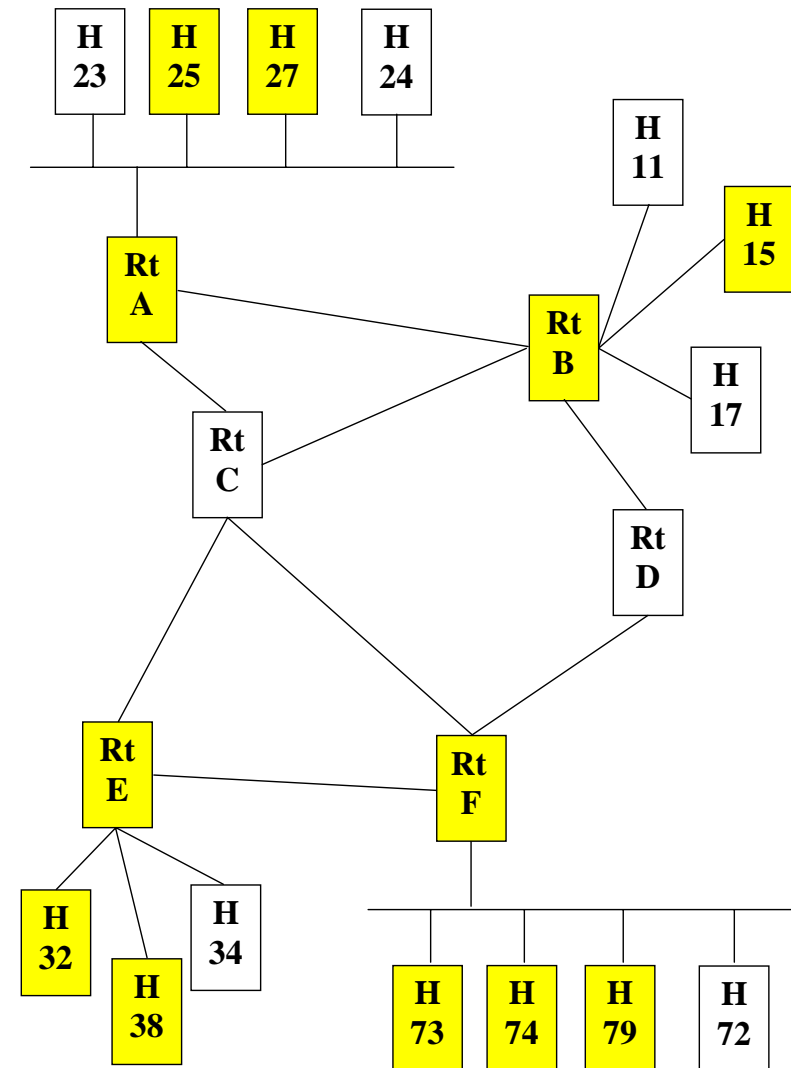
Abstand der Router zu A

	A	B	C	D	E
0:	0	1	2	3	4
1:	-	3	2	3	4
3:	-	3	4	3	4
4:	-	5	4	5	4
5:	-	5	6	5	6
6:	-	7	6	7	6
7:	-	7	8	7	8
8:	-	9	8	9	8

Für B, C, D, und E ist Router A noch erreichbar.

Multicast-Routing:

Die Rechner H25, H27, H15, H73, H74 und H79 mögen eine Multicastgruppe bilden, die beteiligten Router sind RtA, RtB, RtE und RtF.



Bemerkungen zum Multicast-Routing im Internet:

Die Bildung von Kommunikationsgruppen wird in den Internet Group Management Protokollen RFC 1112, RFC 2236 und RFC 3376 geregelt.

Die Mitgliedschaft in einer Gruppe ist dynamisch, jederzeit kann ein Host einer Gruppe beitreten und sie verlassen.

Die Anzahl der Teilnehmer einer Gruppe ist unbegrenzt.

Die Gruppen werden dezentral gebildet.

Ein Host kann an eine Gruppe senden, der er nicht angehört.

Für Multicast-Gruppen ist der Adreßbereich von 224.0.0.0 bis 239.255.255.255 reserviert.

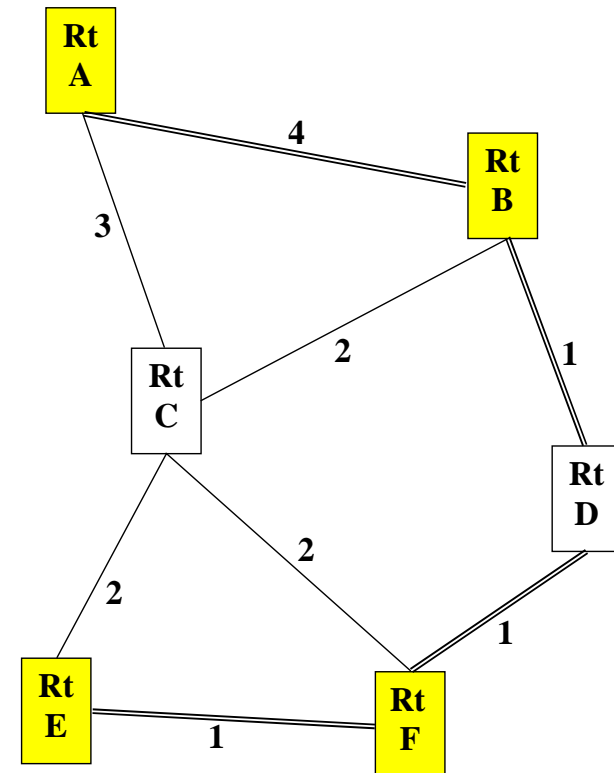
Die Kommunikation zwischen den Teilnehmern einer Multicast-Gruppe, sofern sie bekannt sind, kann über normale Unicast-Nachrichten erfolgen, was sicherlich nicht wünschenswert ist.

Man kann einen allgemeinen Nachrichten-Verteilungs-Baum für eine Routergruppe berechnen. Dies ist eine Variante des Steiner-Baum-Problems, dessen exakte Lösung anerkanntermaßen sehr rechenaufwendig ist. Daher setzt man hier gerne Heuristiken ein.

Die zweite Möglichkeit besteht darin, von jedem Sender einen quellenbasierten Routingbaum zu erstellen. Die Gesamtheit dieser Bäume läßt sich dann vereinfachen.

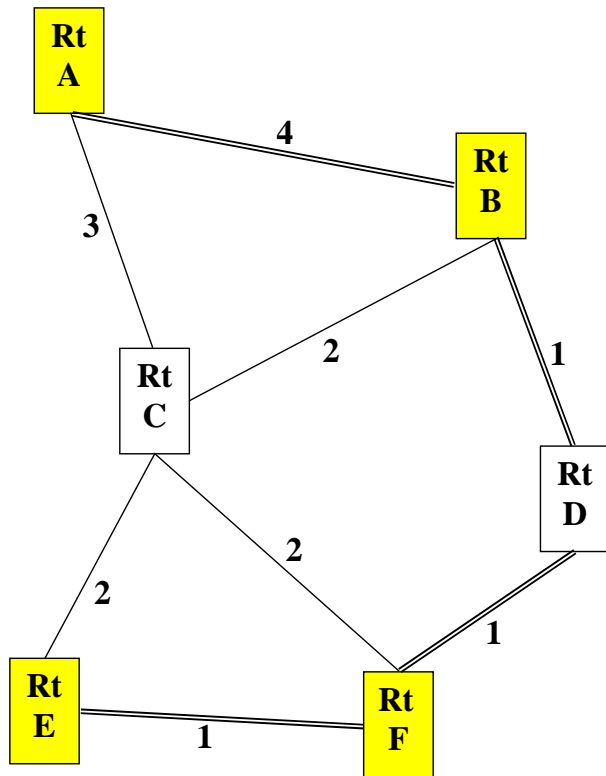
Multicast-Routing:

Berechnung eines gemeinsamen minimalen Gruppenbaums:



Multicast-Routing:

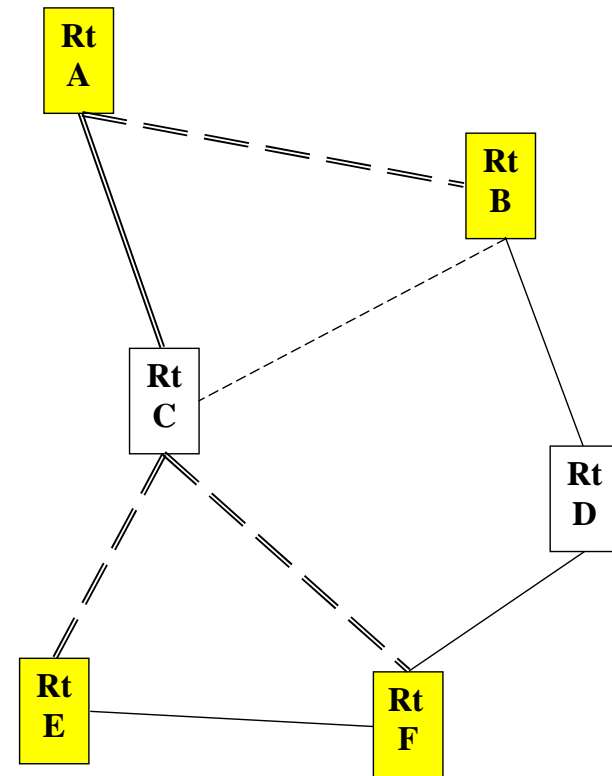
Bildung eines zentrumbasierten Baums:



Als Zentrum wird B gewählt, dem Zentrumsbaum treten E, F und A in dieser Reihenfolge bei.

Multicast-Routing:

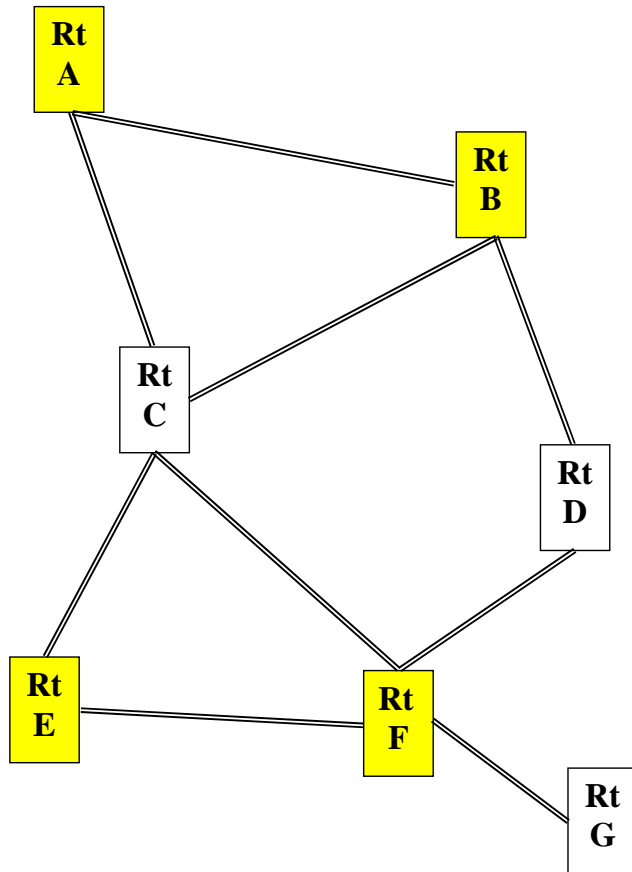
Quellbasierte Routingbäume für A und B:



Bemerkung: In diesem Beispiel wurde eine einheitliche Kantenbewertung von 1 angenommen.

Multicast-Routing: Reverse Path Forwarding:

Prinzip: Ein Router leitet nur Nachrichten weiter, die er über einen kürzesten Pfad zur Quelle erhalten hat. So wird Router C alle Pakete verwerfen, die er von A über B erhält.



Ursprüngliche Struktur des Telefonnetzes / 64-kbs⁻¹-ISDN der Deutschen Telekom:

Auslandskopfvermittlungsstelle:

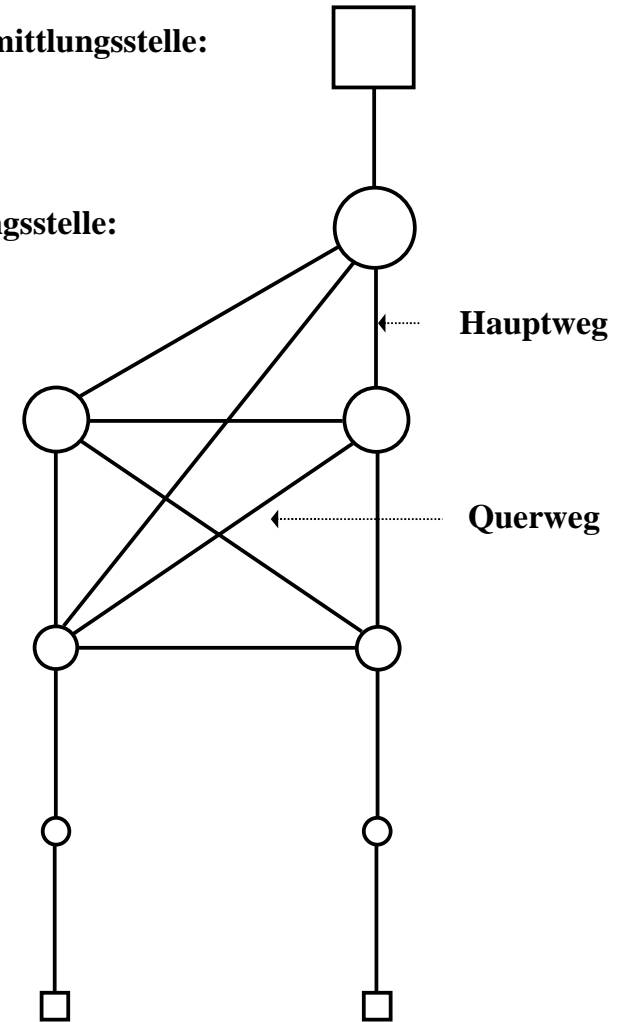
Zentralvermittlungsstelle:

Hauptvermittlungsstelle:

Knotenvermittlungsstelle

Teilnehmervermittlungsstelle

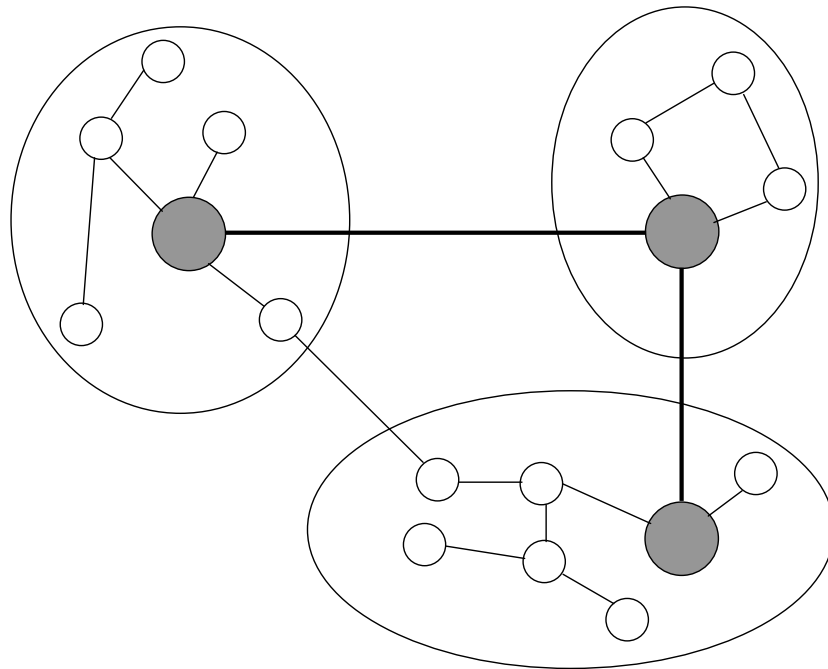
Teilnehmer:



Bemerkung: Die Deutsche Telekom reduziert die Zahl der Hierarchiestufen.

Hierarchisches Routing:

Ein großes Netz wird in Regionen unterteilt; in jeder Region existiert mindestens ein Router, über den der Verkehr mit den anderen Regionen abgewickelt wird.



Bemerkung: Das Netz der ausgezeichneten Router bezeichnet man als "Backbone"-Netz. Die Zweistufigkeit der Routing-Information führt manchmal zu längeren Wegen.

Abschließende Bemerkungen:

- (i) Das Internet Protokoll bietet die Möglichkeit, den vollständigen Weg zum Zielrechner im Paketkopf zu benennen. In diesem Fall spricht man von "Source Routing".
- (ii) Der Algorithmus von Bellman und Ford wird in den "Routing Information" Protokollen eingesetzt. Als Distanzmaß benutzt man die Zahl der zu durchquerenden Router, wobei man ursprünglich die Zahl 16 mit "nicht erreichbar" gleichsetzte. Zur Beschleunigung der manchmal langsamen Konvergenz nutzt man Techniken wie "split horizon" und "poison reverse".
- (iii) Das Protokoll OSPF = "Open Shortest Path First" nutzt in jedem Router Dijkstras Algorithmus. Zuvor müssen mittels eines verlässlichen Flutprozesses die lokalen Informationen eines jeden Routers netzweit verbreitet werden. Für die Bedeutung von "open" werden allgemein zwei Erklärungen angeboten.
- (iv) Im Internet setzt man das "Border Gateway Protocol" ein, um eine konsistente Sicht des aus autonomen Untersystemen bestehenden Internet zu erlangen. BGP nutzt seinerseits TCP.