Routing im Internet

chaos seminar

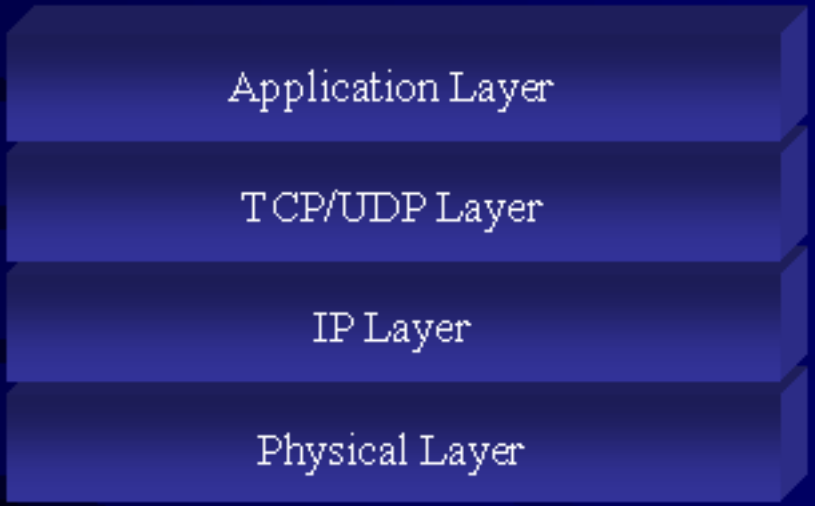
6.9.99

frank@kargl.net

Routing im Internet

- Was ist Routing?
- Direct vs. Indirect Delivery
- Routing Tabellen (Hosts vs. Router)
- Interior Routing Protokolle
  - Distance Vector (RIP)
  - Link State (OSPF)
- Exterior Routing Protokolle
  - BGP
- Aufbau und Organisation des Routing im Internet

TCP/IP Protokoll-Stack



Host

Host + Network

Aufbau einer TCP-Verbindung

134.60.1.14

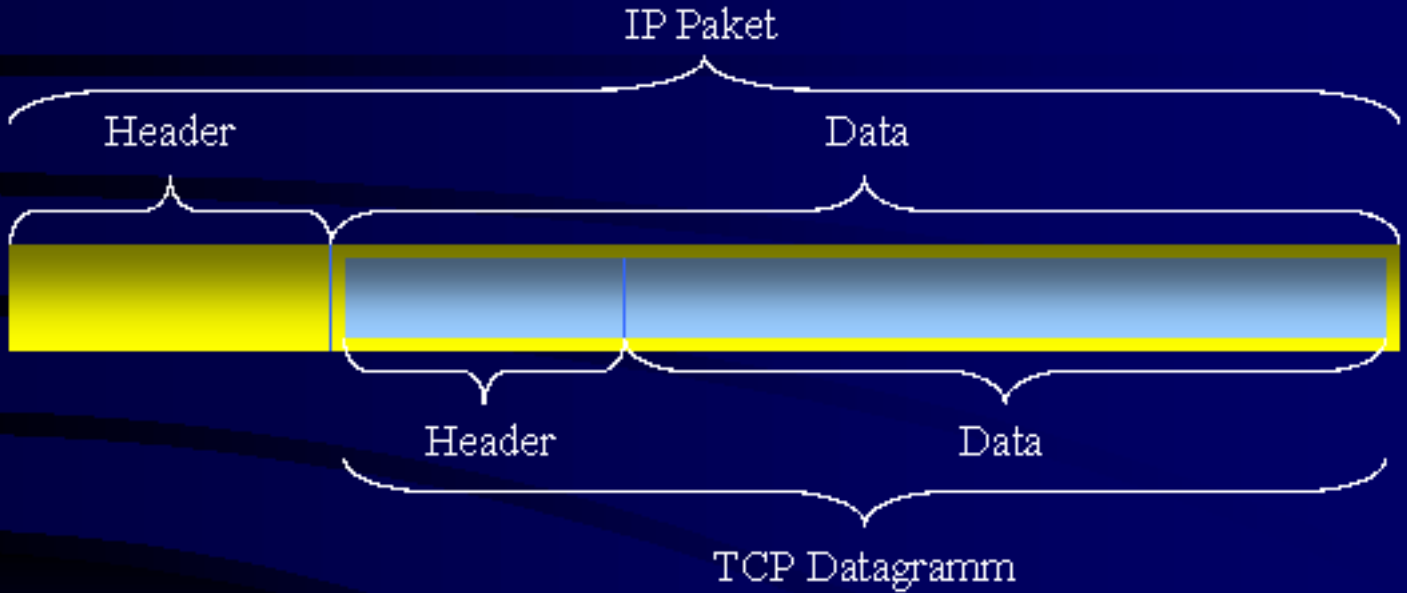
194.25.243.1

Daten

Port 36732

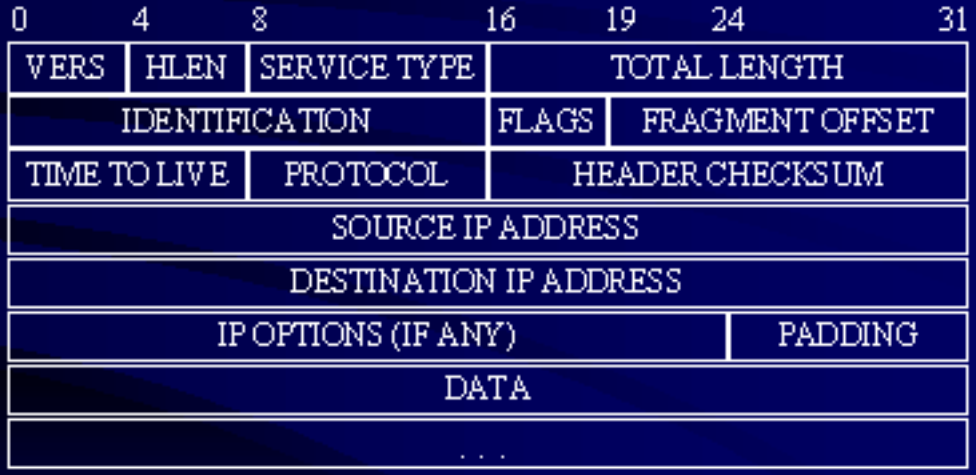
Port 80

TCP/IP Paketaufbau



IP Header:	Source + Destination Address
TCP Header:	Source + Destination Port
TCP Data:	Application Protocol (z.B. http)

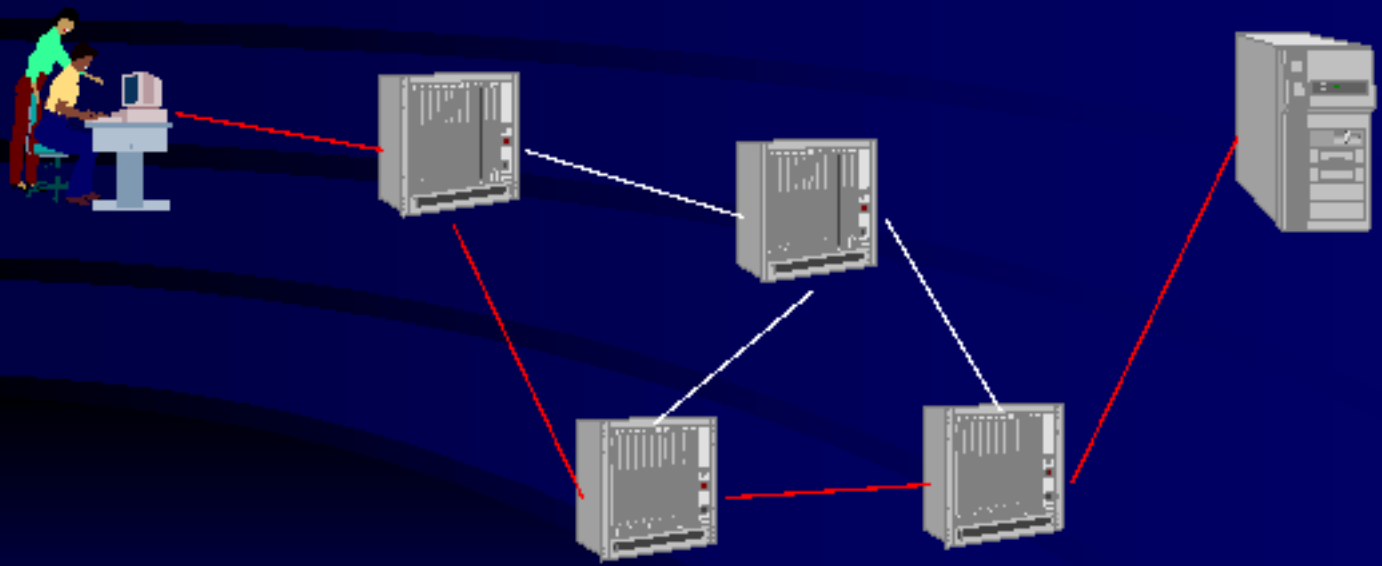
Aufbau IP Paket



IP-Routing

134.60.1.14

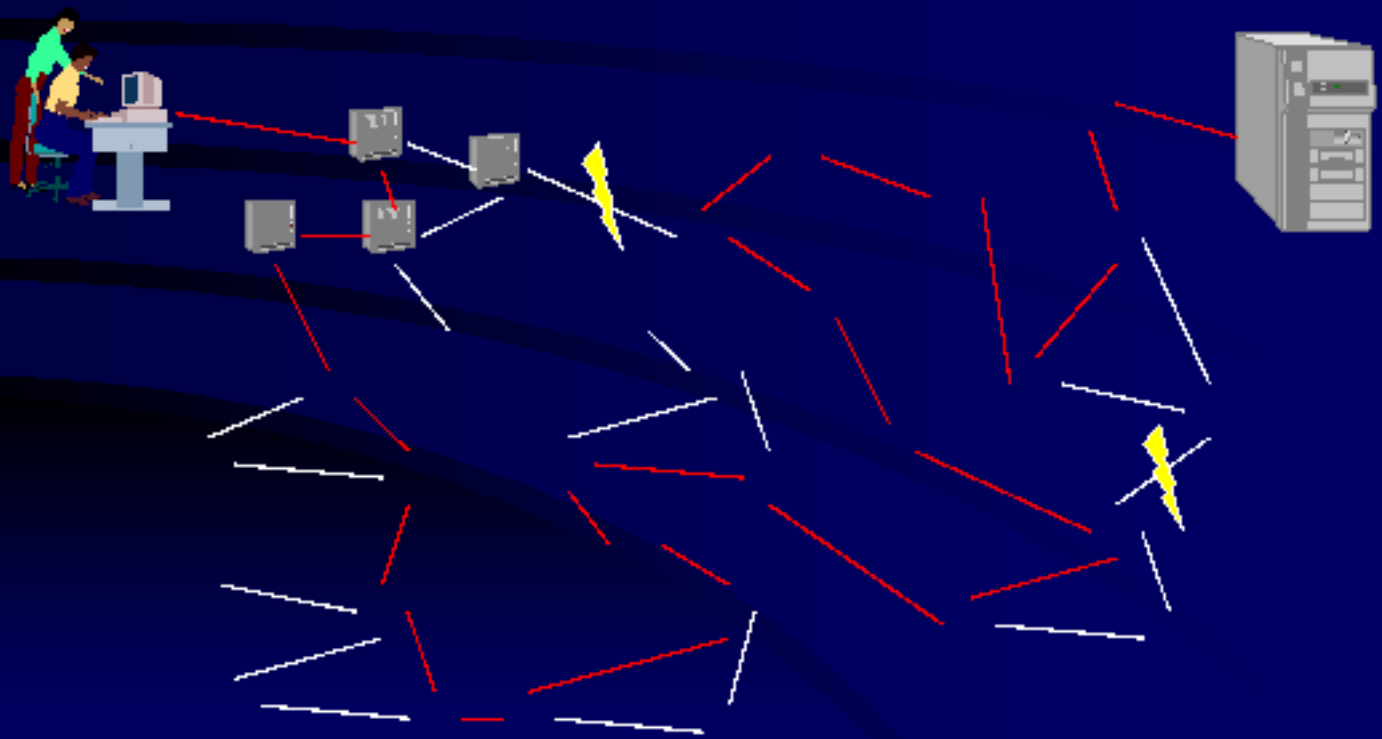
194.35.243.1



IP-Routing

134.60.1.14

194.35.243.1



IP-Routing

Routing:

Wegewahl anhand der gegebenen Zieladresse

Randbedingungen:

- Delays
- Link-State
- Kosten
- Policies uvm.

Direct vs. Indirect Deliveriy

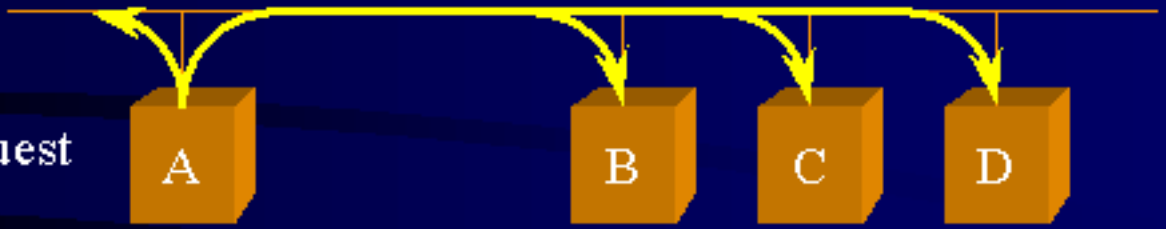
- Direct delivery
  - Source und Destination im gleichen Netzwerk
- Indirect delivery
  - Source und Destination nicht im gleichen Netzwerk

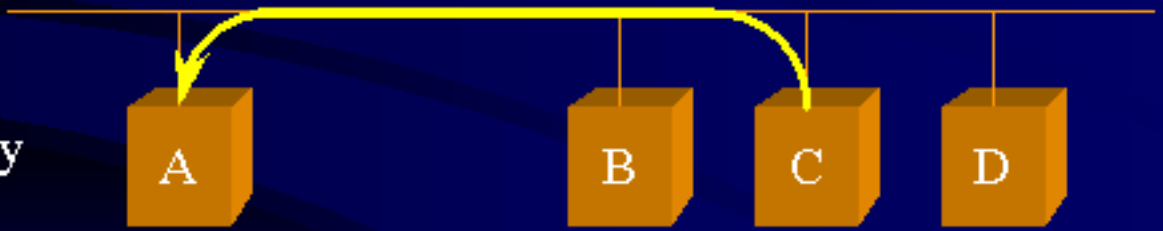
Direct Delivery

Direct Delivery

- Wenn Direct Delivery möglich, direktes Senden über physikalisches Netzwerk
- Physikalische Netzwerkadresse bestimmen (via ARP)

ARP





ARP

- ARP sorgt für dynamisches Zuordnung von IP- zu physikalischen Adressen.
- Normalerweise mittels Broadcast
- Tuning:
  - local caching vermeidet unnötige Anfragen
  - Jeder ARP Broadcast enthält die IP- und physikalische Adresse des Senders

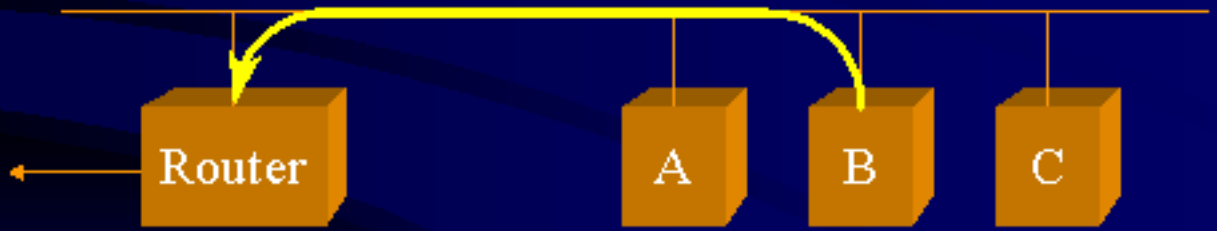
Indirect Delivery

- Zielrechner nicht im eigenen Subnetz => Auslieferung über Router
- Nächster Router (next hop) im eigenen Subnetz
- aus Routing Tabelle
- Zustellung an Router via Direct Delivery

Routing Tabellen

Bei (normalen) Hosts:

- Manuell konfiguriert
- Default-Router für Verkehr außerhalb des eigenen Subnet




```
scuba:~ # netstat -rn
Kernel IP routing table
Destination     Gateway         Genmask         Flags   MSS Window  irtt Iface
134.60.77.0     0.0.0.0         255.255.255.0   U         0 0          0 eth0
127.0.0.0       0.0.0.0         255.0.0.0       U         0 0          0 lo
0.0.0.0         134.60.77.99    0.0.0.0         UG        0 0          0 eth0
```

Routing Tabellen

- Woher wissen Router (oder Hosts mit mehreren Interfaces) den "next hop" ?


```
[fkargl@wega fkargl]$ netstat -rn
Routing Table:
  Destination           Gateway           Flags  Ref   Use   Interface
------------------- ------------------- ----- ----- ------ ---------
134.60.222.160      134.60.240.102       UGH      0      0
134.60.222.161      134.60.240.102       UGH      0      0
134.60.7.143        134.60.240.102       UGH      0      0
134.60.26.0         134.60.240.43        UG       0      6
134.60.122.0        134.60.240.45        UG       0      0
134.60.90.0         134.60.240.103       UG       0      0
134.60.218.0        134.60.1.25          UG       0      0
... many more ...
```

Routing Protokolle

- Router informieren sich gegenseitig über die Netztopologie
- Distance Vector Protokolle (RIP)
- Link State Protokolle (OSPF)

Distance Vector

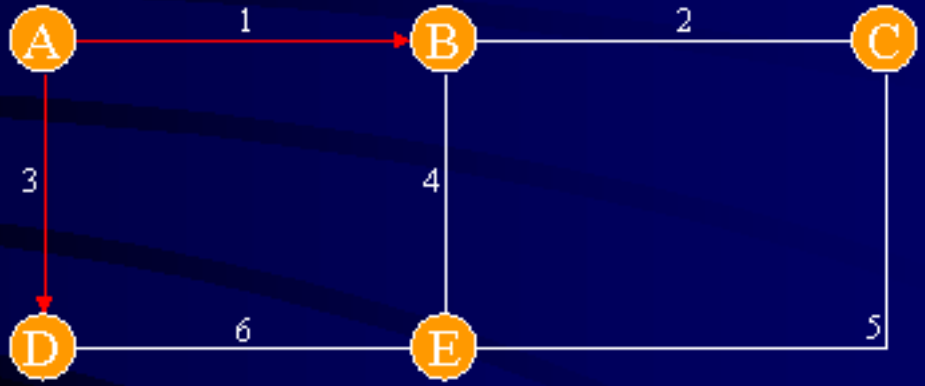
- basieren auf Bellmann-Ford Algorithmus
- verteiltes Berechnen der Topologie
- Zyklischer Austausch von Distanz-Vektoren

Distance Vector

Von A nach	Link	Cost
A	local	0

Von B nach	Link	Cost
A	1	1
B	**local**	**0**

Von C nach	Link	Cost
C	local	0



Von D nach	Link	Cost
A	3	1
D	**local**	**0**

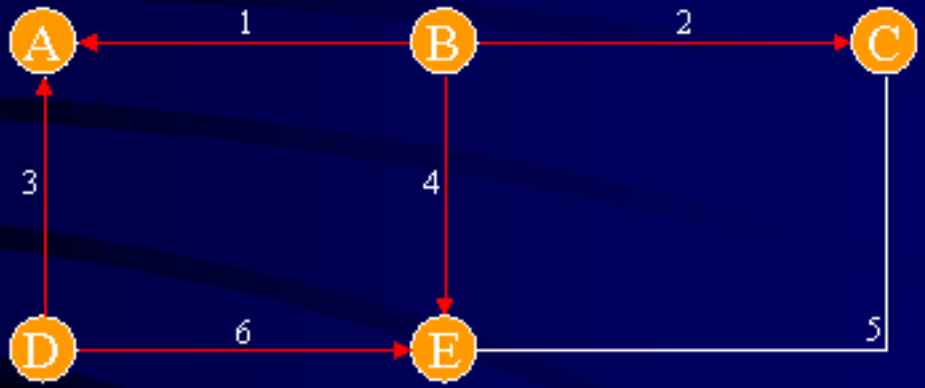
Von E nach	Link	Cost
E	local	0

Distance Vector

Von A nach	Link	Cost
A	local	0
B	**1**	**1**
D	**3**	**1**

Von B nach	Link	Cost
A	1	1
B	local	0

Von C nach	Link	Cost
A	**2**	**2**
B	**2**	**1**
C	local	0



Von D nach	Link	Cost
A	3	1
D	local	0

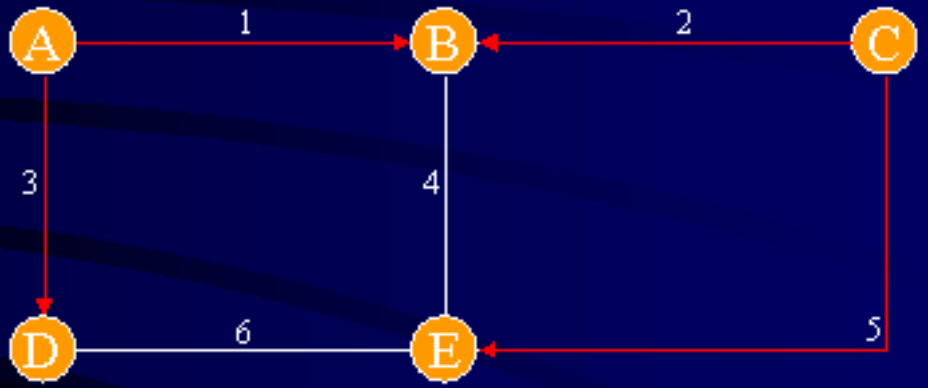
Von E nach	Link	Cost
A	**4**	**2**
A	**6**	**2**
B	**4**	**1**
D	**6**	**1**
E	local	0

Distance Vector

Von A nach	Link	Cost
A	local	0
B	1	1
D	3	1

Von B nach	Link	Cost
A	1	1
B	local	0
C	**2**	**1**
D	**1**	**2**

Von C nach	Link	Cost
A	2	2
B	2	1
C	local	0



Von D nach	Link	Cost
A	3	1
B	**3**	**2**
D	local	0

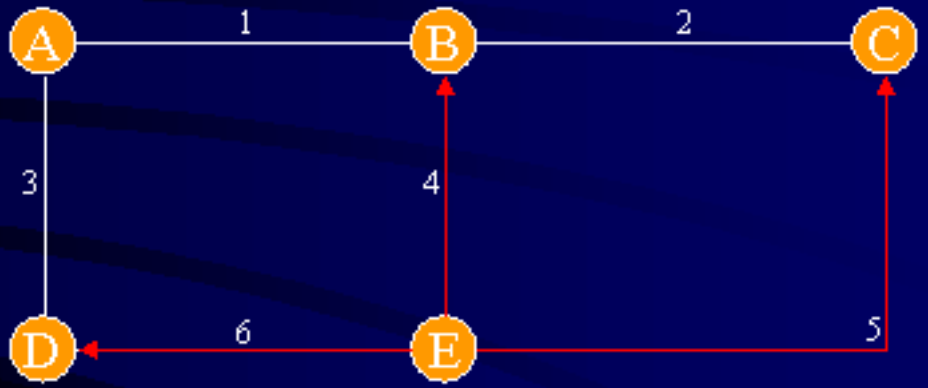
Von E nach	Link	Cost
~~A~~	~~2~~	~~3~~
A	4	2
~~A~~	~~6~~	~~2~~
B	4	1
~~B~~	~~5~~	~~2~~
C	**5**	**1**
D	6	1
E	local	0

Distance Vector

Von A nach	Link	Cost
A	local	0
B	1	1
D	3	1

Von B nach	Link	Cost
A	1	1
B	local	0
C	2	1
D	1	2
E	**4**	**1**

Von C nach	Link	Cost
A	2	2
B	2	1
C	local	0
E	**5**	**1**



Von D nach	Link	Cost
A	3	1
B	3	2
C	6	2
D	local	0
E	**6**	**1**

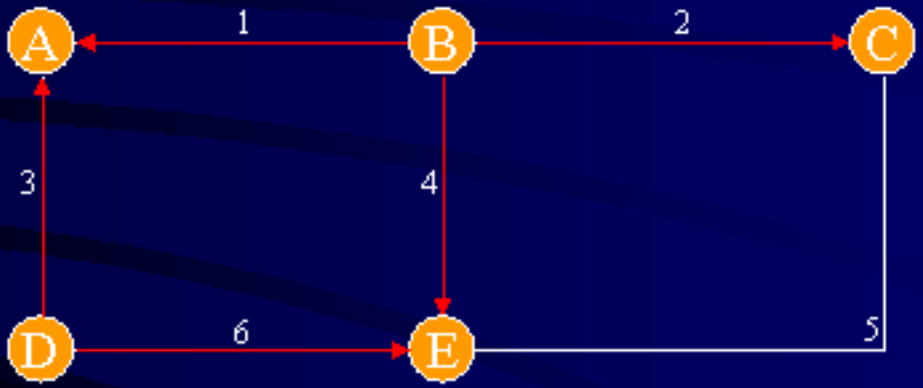
Von E nach	Link	Cost
A	4	2
B	4	1
C	5	1
D	6	1
E	local	0

Distance Vector

Von A nach	Link	Cost
A	local	0
B	1	1
C	**1**	**2**
D	3	1
E	**4**	**2**

Von B nach	Link	Cost
A	1	1
B	local	0
C	2	1
D	1	2
E	4	1

Von C nach	Link	Cost
A	2	2
B	2	1
C	local	0
D	**2**	***3***
E	5	1



Von D nach	Link	Cost
A	3	1
B	3	2
C	6	2
D	local	0
E	6	1

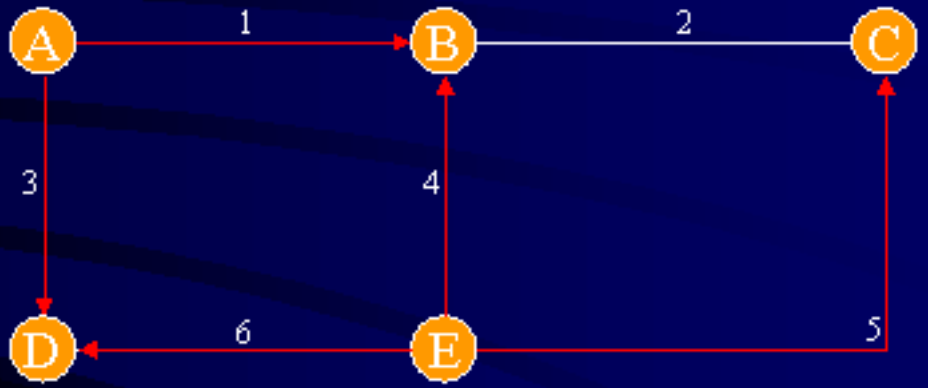
Von E nach	Link	Cost
A	4	2
B	4	1
C	5	1
D	6	1
E	local	0

Distance Vector

Von A nach	Link	Cost
A	local	0
B	1	1
C	1	2
D	3	1
E	4	2

Von B nach	Link	Cost
A	1	1
B	local	0
C	2	1
D	1	2
E	4	1

Von C nach	Link	Cost
A	2	2
B	2	1
C	local	0
D	**5**	**2**
E	5	1



Von D nach	Link	Cost
A	3	1
B	3	2
C	6	2
D	local	0
E	6	1

Von E nach	Link	Cost
A	4	2
B	4	1
C	5	1
D	6	1
E	local	0

Distance Vector

- Höchster Wert = Infinity = nicht erreichbar
- Probleme bei Link Ausfällen:
  - Routing Loops (Bouncing Effekt)
  - langsame Konvergenz (Counting to Infinity)
- Lösungsansätze
  - Split Horizon
  - Triggered Updates

RIP V1

- RFC 1058, Juni 1988
- BSD Unix
- Metric "Hop Count"
- einfaches UDP Protokoll

RIP V2

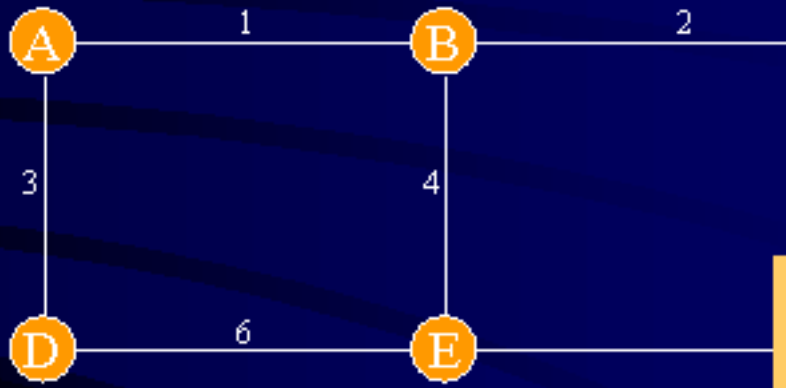
Verbesserungen:

- Authentisierung
- Subnetzmaske
- Multicast statt Broadcast
- Routing Domains
- Rückwärtskompatibel

Link State

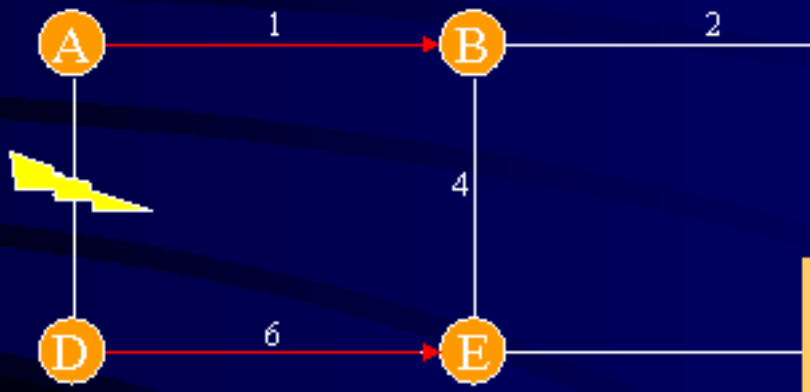
- Komplette Netztopologie auf allen Routern bekannt
- Berechnung optimaler Routen nach Shortest Path First Algorithmus von Dijkstra
- Komponenten:
  - Link State Database
  - Flooding Protocol

Link State Database



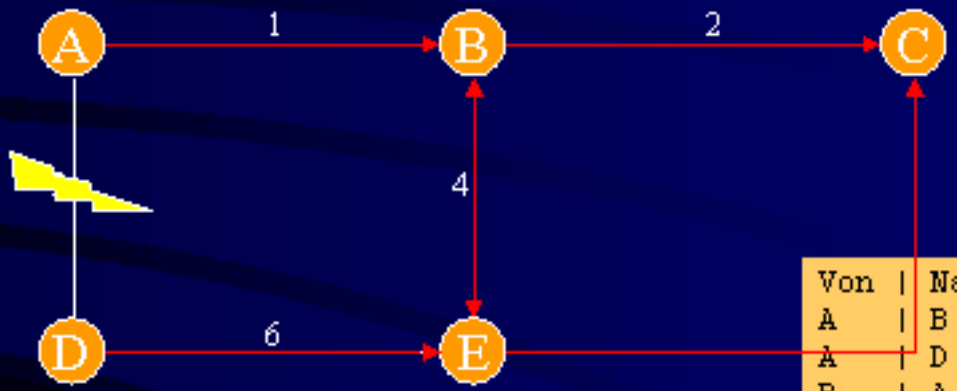
Von	Nach	Link	Dist.
A	B	1	1
A	D	3	1
B	A	1	1
B	C	2	1
B	E	4	1
C	B	2	1
C	E	5	1
D	A	3	1
D	E	6	1
E	B	4	1
E	C	5	1
E	D	6	1

Link State Database



Von	Nach	Link	Dist.
A	B	1	1
A	D	3	inf.
B	A	1	1
B	C	2	1
B	E	4	1
C	B	2	1
C	E	5	1
D	A	3	inf.
D	E	6	1
E	B	4	1
E	C	5	1
E	D	6	1

Link State Database



Von	Nach	Link	Dist.
A	B	1	1
A	D	3	**inf.**
B	A	1	1
B	C	2	1
B	E	4	1
C	B	2	1
C	E	5	1
D	A	3	**inf.**
D	E	6	1
E	B	4	1
E	C	5	1
E	D	6	1

OSPF

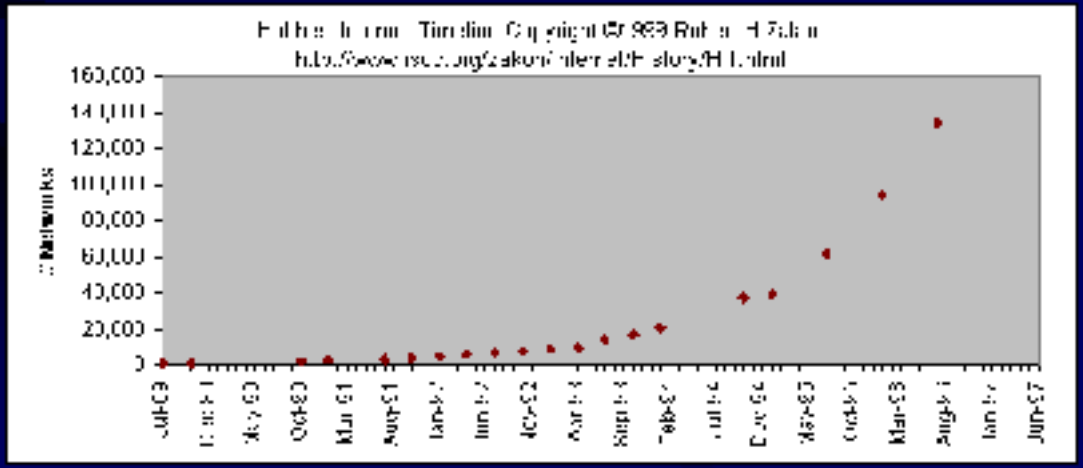
- Open SPF
- Vorteile gegenüber RIP
  - Konvergiert schneller
  - keine Routing Loops
  - mehrere Metriken parallel
  - Lastverteilung auf mehrere Pfade
  - Bessere Verwaltung größerer Netzwerke
  - weniger Netzverkehr

OSPF Features

- Broadcast / Non-Broadcast Networks
- Multiple Areas (insbes. Backbone)
- Stub Areas
- Bestandteile:
  - Hello Protocol
  - Exchange Protocol
  - Flooding Protocol

Exterior Gateway Protocols

- Bisherige Protokolle für Routing innerhalb eines LANs
- Problem im Internet:



Exterior Gateway Protocols

- Lösung: Einführung sog. Autonomous Systems (AS)
- BelWue AS553
- AS verbirgt die interne Struktur, nur noch Netzwerk-Adressen werden weitergegeben
- AS verantwortlich für das Routing innerhalb der AS (z.B. via OSPF)
- Border Gateways an AS-Übergängen

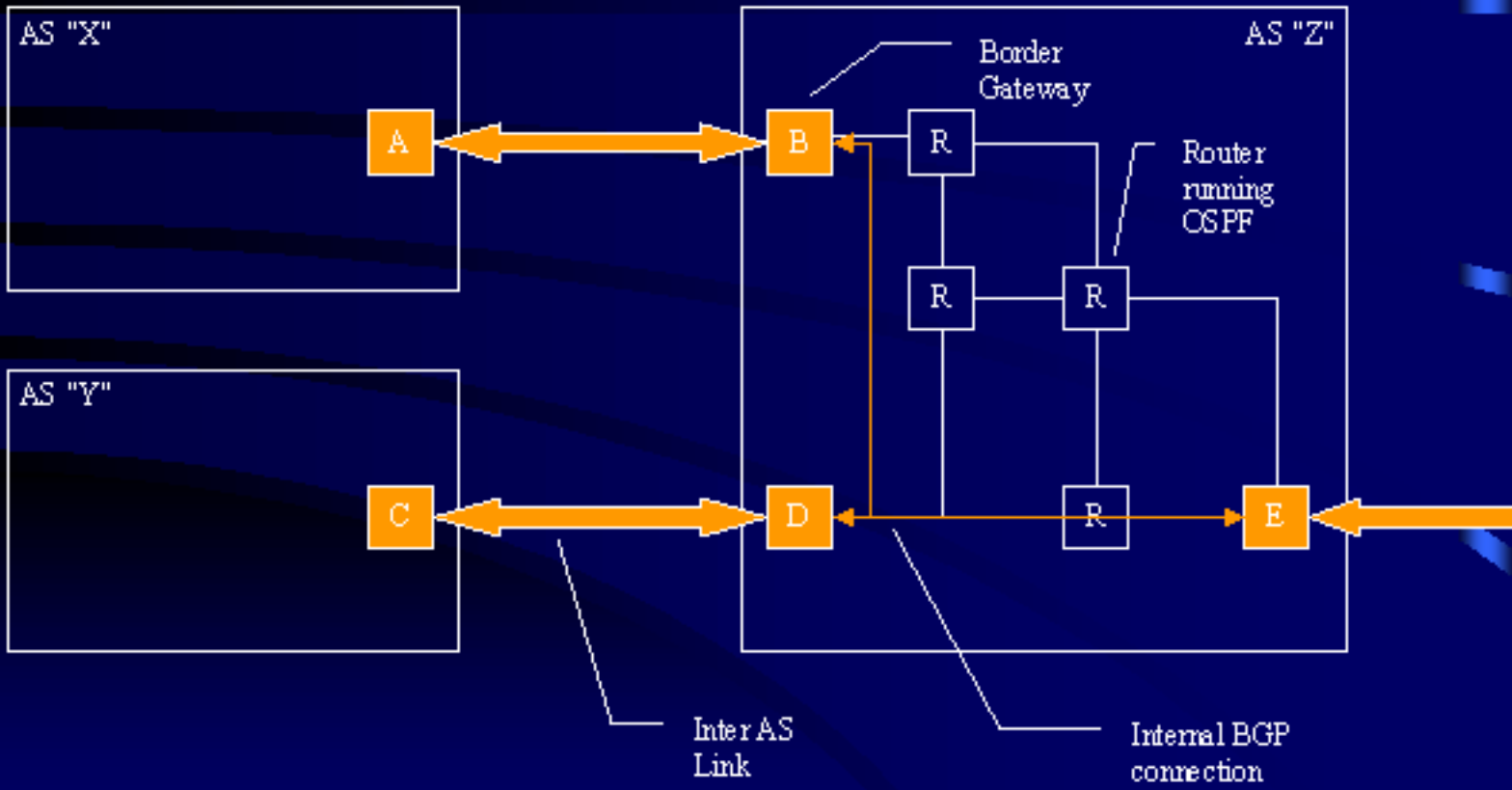
Exterior Gateway Protocols

- EGP bis Ende der 80er
- Nachteile:
  - Baumstruktur
  - Kein Schutz vor falschen Informationen
  - Schlechte Unterstützung für Policy Routing
  - Schlechte Skalierbarkeit

Border Gateway Protocol

- Heute verwendetes EGP
- RFC-1105/1163/1267/1771 (BGP 1/2/3/4)
- in Arbeit: BGP 4+ (IPv6)
- Path Vector Prinzip
- Routing Information enthält AS# aller bisher durchlaufenen AS => Loop Detection
- freie Path Attributes => Policy Routing
- Processing ähnlich Distance Vector

Border Gateway Protocol



Border Gateway Protocol

Policy Routing:

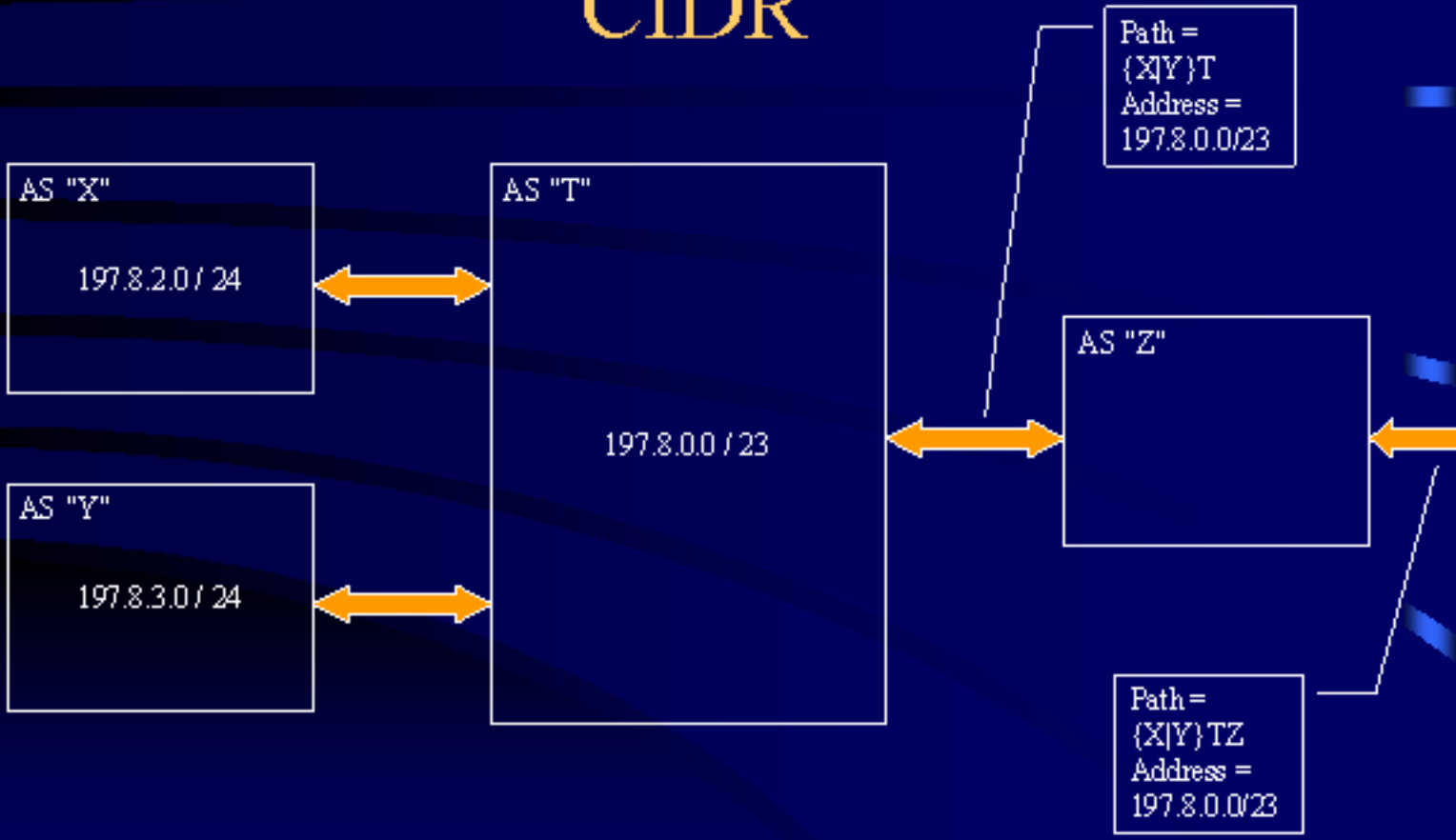
- Acceptable Use Policy
- Number of AS, AS Weights, Block AS, Stability
- Policy freies Routing

Border Gateway Protocol

CIDR:

- Classless Interdomain Routing
- BGP 4
- Lösung für:
  - Class B Exhaustion (16k, bis März '94)
  - Routing Table Explosion (#Dest*#Neighbors) 100MByte und mehr
- Contigous Class C Networks
- Routing Table Aggregation (same prefix)

CIDR


Path =
{X|Y}T
Address =
197.8.0.0/23
AS "X"
197.8.2.0 / 24
AS "T"
197.8.0.0 / 23
AS "Z"
AS "Y"
197.8.3.0 / 24
Path =
{X|Y}TZ
Address =
197.8.0.0/23

Class C Verteilung

Multiregional:	192.0.0.0 - 193.255.255.255
Europa:	194.0.0.0 - 195.255.255.255
Others:	196.0.0.0 - 197.255.255.255
North America:	198.0.0.0 - 199.255.255.255
Central/S.America:	200.0.0.0 - 201.255.255.255
Pacific Rim:	202.0.0.0 - 203.255.255.255
Others:	204.0.0.0 - 205.255.255.255
Others:	206.0.0.0 - 207.255.255.255

(geplant)

- Geographisch vs. Provider-based ?

Vergleich

IGP:

- möglichst automatisches Management kleinerer Systeme
- Routing ergibt sich aus Topologie
- schnellstmögliche Konvergenz und Routenberechnung

EGP:

- Verwaltet das Zusammenspiel vieler unabhängiger Teilnetze
- Routing wird maßgeblich durch das Mangement festgelegt (Policy)
- Dient hauptsächlich dem automatischen Umschalten auf Backup Routen

What else?

- Multicast Routing (M-OSPF, PIM)
- Mobile IP
- Resource Reservation
- IPv6
- IANA, ICANN, Internic, RIPE