

Datenkommunikation

Teil 3.2: Internetschicht: Routing

O.Univ.Prof.Dr. Harmen R. van As

Übersicht

3.3a Internet-Referenzmodell: Routing

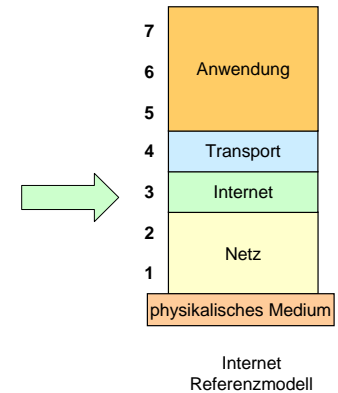
- Routingverfahren
- Netzaufteilung

Austausch von Routing Informationen

- OSPF (Open Shortest Path First)
- BGP (Border Gateway Protocol)

Routenbestimmung

- Dijkstra Algorithmus



Routing

statisch

- Routing-Tabellen werden manuell gesetzt
- fehleranfällig
- einfach für kleine Netze

dynamisch

- Routing-Tabellen werden durch Austausch von Routing-Protokollnachrichten gelernt
- Aufsetzen von gleichen Routing-Protokollen in einer Domäne

Protokolle und Routing im Internet

Protokolle der Vermittlungsschicht

- IP (Internet Protocol)
- ARP (Address Resolution Protocol)
- RARP (Reverse ARP)
- ICMP (Internet Control Message Protocol)
- IGMP (Internet Group Management Protocol)
- SNAP (Subnetwork Access Protocol)

Routingprotokolle

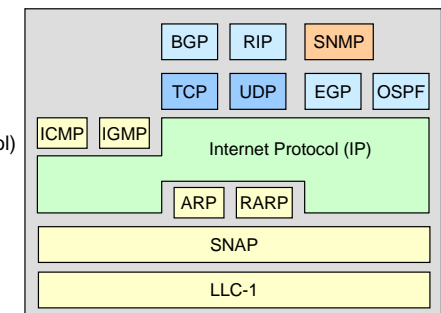
- RIP (Routing Information Protocol)
- BGP (Border Gateway Protocol)
- EGP (External Gateway Protocol)
- OSPF (Open Shortest Path First)

Transportprotokolle:

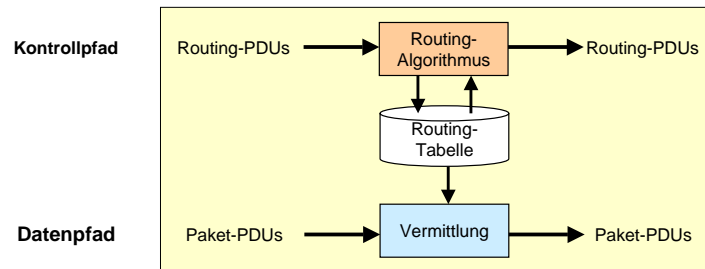
- UDP (Universal Datagram Protocol)
- TCP (Transmission Control Protocol)

Netzmanagement

- SNMP (Systems Network Management Protocol)



Kontroll- und Datenpfad in Routern



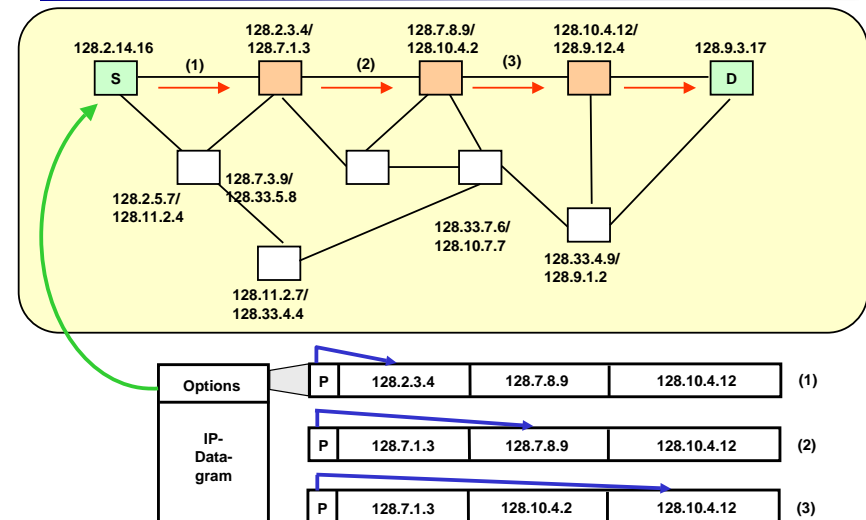
Kontrollpfad

- Steuert das Ruten der Daten, ist aber nicht direkt im Routing-Prozess involviert
- Routingprotokolle sind oberhalb der Schicht 3 angesiedelt
- Die Aktualisierung der Routingtabelle geschieht durch den jeweils eingesetzten Algorithmus
- Routingtabelle enthält Routinginformation, die das Weiterleiten der Pakete ermöglicht
- Wegewahl beim Routing wird anhand der Routinginformation in der Routingtabelle durchgeführt

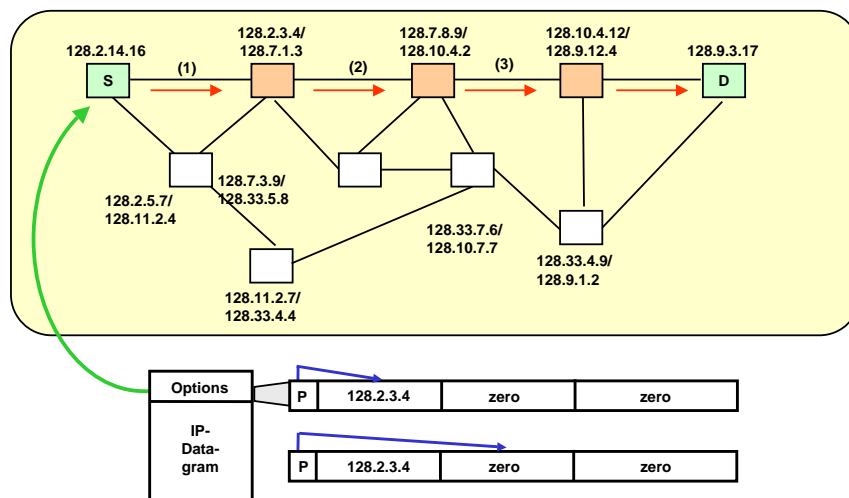
Datenpfad

Vermittlung der Pakete auf Schicht 3 (Vermittlungsschicht)

IP Source Routing



IP Record Routing



Dynamik von Routingverfahren

Wie dynamisch ist das Routingverfahren?

Nicht adaptiv

- (Routen ändern sich nur sehr selten)
- Oft durch manuelle Änderungen Routenänderungen sind seltener als Verkehrsänderungen

Adaptiv

- (Routen ändern sich in Abhängigkeit des Verkehrs bzw. der Netztopologie)
- Aktueller Zustand des Netzes wird damit berücksichtigt
- Schleifen und Oszillationen in Routen wahrscheinlicher als bei nicht adaptiven Verfahren

Können **periodisch** operieren oder in direkter **Reaktion auf Änderungen**

Zielkonflikt:

- Knoten haben veraltete oder unvollständige Informationen über den Zustand des Netzes
- Evtl. hohe Belastung durch Austausch von Routinginformationen

Routingprotokolle

Im praktischen Einsatz: verteilte adaptive Routingalgorithmen
Dabei werden zwei grundlegende Algorithmen unterschieden

Distanz-Vektor-Algorithmen

- Distanz ist Routing-Metrik
- jedes System kennt die Distanz zu allen anderen Systemen im Netz
hierzu werden die aktuellen Distanzen zwischen den Nachbarn ausgetauscht

Problem

- kürzerer langsamerer Weg wird längerem schnelleren Weg vorgezogen

Beispiele: Routing Information Protocol (RIP), Distance-Vector-Routing-Protocol (DVRP)

Link-State-Algorithmen

- Unterschiedliche Routing-Metriken möglich
- berücksichtigt die aktuellen Zustände der Netzanschlüsse
- jeder Router kennt die komplette Netztopologie und berechnet seine Routinginformation
- Link-State-Algorithmen konvergieren meistens schneller als Distanz-Vektor-Algorithmen
- für größere Netze sind sie potenziell besser geeignet

Beispiele: Open Shortest Path First (OSPF)

Intra-Domain Intermediate System to Intermediate System Routing Protocol (IS-IS)

Link-State-Routing

Grundlegende Vorgehensweise

- Knoten müssen am Anfang nur ihre direkten Nachbarn kennen
- Entdecken neuer Nachbarn mit speziellen Dateneinheiten (z.B. HELLO)
- Bestimmen der Kosten zu den direkten Nachbarn

Link State Broadcast

- Identität und Kosten zu den direkten Nachbarn werden an alle Knoten im Netz durch Fluten weitergeleitet (Broadcast)
- Knoten können Topologie lernen durch die Link State Broadcasts der anderen Knoten

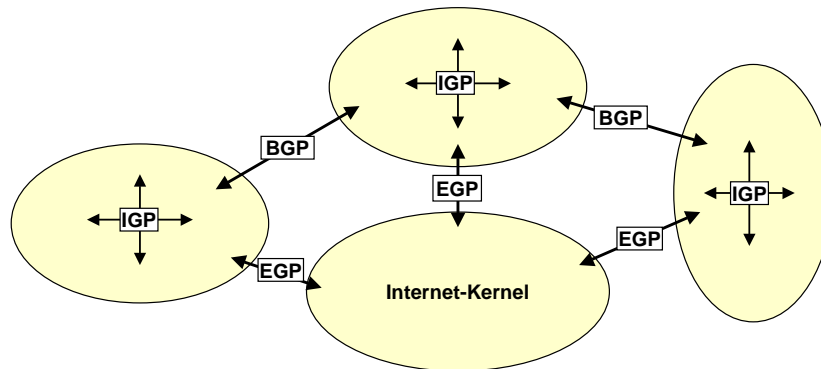
Ergebnis:

- Alle Knoten haben identisches Wissen über das Netz
- Berechnung der kürzesten Pfade durch Link-State-Algorithmus
- Jeder Knoten berechnet die kürzesten Pfade
- Die berechneten Pfade sind aufgrund der identischen Information gleich
- Nach Fluten und Berechnung der kürzesten Pfade in jedem Knoten ist das Netz schleifenfrei und in stabilen Zustand konvergiert

Dijkstra Algorithmus

- Berechnet den Pfad mit den geringsten Kosten von einem Knoten zu allen anderen Knoten im Netz

Routing Protocols



Interior Gateway Protocol (IGP)
Enhanced IGP
Exterior Gateway Protocol (EGP)
Border Gateway Protocol (BGP)

Routing Information Protocol (RIP)
Open Shortest Path First (OSPF)

Netzaufteilung in Autonome Systeme (AS)

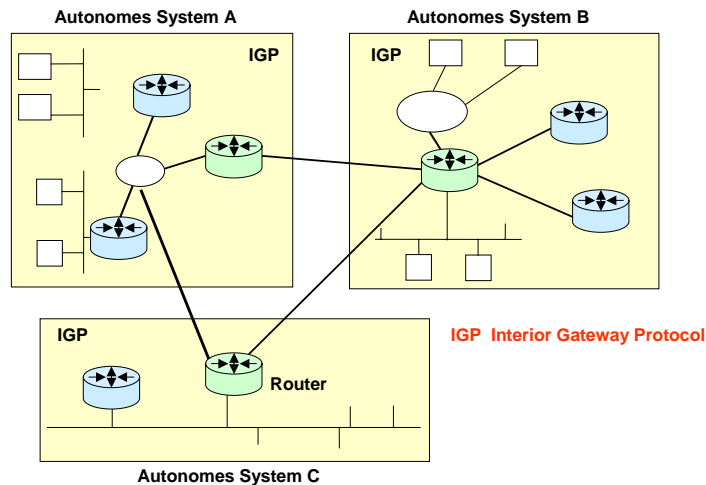
Aufteilung großer Netze in Autonome Systeme (AS) oder Regionen

- **Grund:** Anzahl der Einträge in der Routingtabelle und Menge der ausgetauschten Routinginformation sonst nicht skalierbar mit Netzgröße
- Die Router haben in einem autonomen System normalerweise nur Routing-Informationen über dieses autonome System.
- In jedem autonomen System gibt es zumindest ein ausgezeichnetes Zwischensystem, das als Schnittstelle zu anderen autonomen Systemen dient.

Vorteile

- Skalierbarkeit
- Größe der Routingtabellen ist abhängig von der Größe des autonomen Systems.
- Änderungen von Einträgen in den Routingtabellen werden nur innerhalb eines autonomen Systems weitergegeben.
- Autonomie, Internet = Netz von Netzen
- Routing kann im eigenen Netz kontrolliert werden
- Im administrativen System gibt es ein einheitliches Routingprotokoll
- Routingprotokolle der autonomen Systeme müssen nicht identisch sein

Autonome Systeme (AS)



Routing und Autonome Systeme

Das globale Internet besteht aus **Autonomen Systemen** (Autonomous Systems, AS)

- Jedes AS hat eine eindeutige Nummer (derzeit 16 Bit, Erweiterung auf 32 Bit geplant)

Verbund von Autonomen Systemen

Stub AS

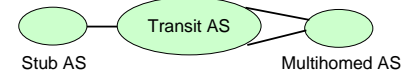
- Kleine Unternehmen
- Anschluss an genau einen Provider

Multihomed AS

- Große Unternehmen
- Anschluss an mehrere Provider
- Kein Transitverkehr

Transit AS

- Provider



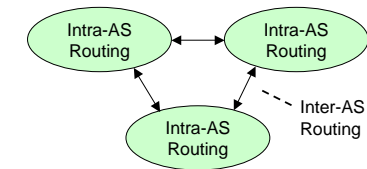
Zwei Ebenen des Routing

Intra-AS

- Administrator ist verantwortlich für Wahl des Routingprotokolls

Inter-AS

- Einheitliche Standards



Intra- und Inter-AS-Routingprotokolle

Policy

Politische Frage: welcher Transit-Verkehr darf das AS passieren?

Inter: Policies werden vom Provider aufgestellt

Intra: Es gibt nur eine Organisation und deshalb sind wenig Policies erforderlich

Skalierbarkeit

Inter: weitere Ebene der Abstraktion ist dadurch gegeben
Tabellengrößen und Anzahl der Updates können reduziert werden, da Ausfälle innerhalb eines AS meistens verborgen bleiben können

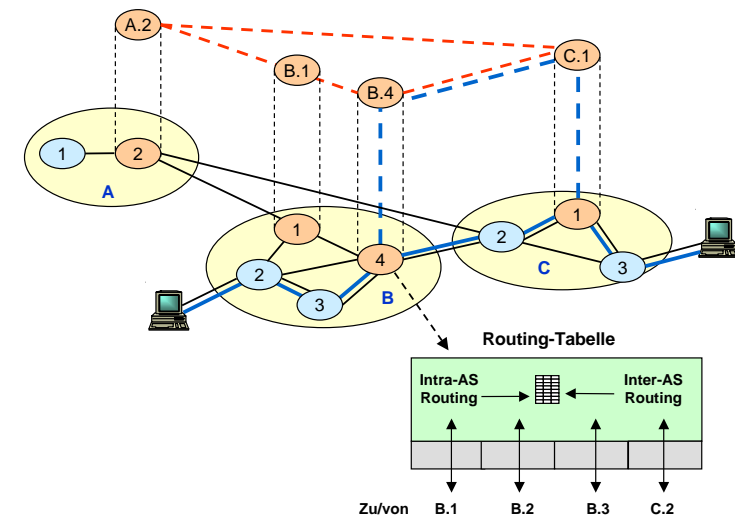
Intra: bessere Stabilität

Leistungsfähigkeit

Inter: Policies sind erforderlich und wichtiger als Leistungs-Metriken

Intra: Konzentration auf Leistungs-Metriken

Inter-AS und Intra-AS-Routing



Intra-AS-Routing

Intra-AS-Routingprotokolle bezeichnet man auch als Interior Gateway Protocols (IGP).

Die bekanntesten Protokolle hierfür sind

- **RIP** (Routing Information Protocol)
- **OSPF** (Open Shortest Path First)
- **IS-IS** (Intra-Domain Intermediate System to Intermediate System Routing Protocol) ursprünglich ISO/OSI-Routingprotokoll, für IP eingesetzt bei großen Providern
- **EIGRP** (Enhanced Interior Gateway Routing Protocol) CISCO proprietär

ICMP Router Discovery Protocol

- **ICMP Router Discovery Protocol (IRDP):**
router advertisement and router solicitation messages to discover IP addresses of routers on directly attached subnets
- **Each router:**
periodically multicast router advertisement messages
- **Hosts:**
Discover addresses of routers by listening for these router advertisement messages
Router solicitation messages to request immediate advertisements
- Not require hosts to process routing protocols
- Not require manual configuration by administrator
- No path optimisation
- Hosts receive redirect messages

OSPF: Open Shortest Path First

- Common header
- Hello packet
- Database description packet
- Link state request packet
- Link state update packet
- Link state acknowledgement packet
- Link state header
- Router links advertisement
- Network links advertisement
- Summary links advertisement
- External links advertisement

Pakete:

- Hello
- Link State Request
- Link State Update
- Link State Acknowledgement

Funktionen:

- Discovering neighbors
- Electing the designated router
- Initializing neighbors
- Propagating link state information
- Calculating route tables

Protokoll OSPF

Konzept: jeder Knoten verfügt über komplette Information der Netztopologie

Link-State-Algorithmus

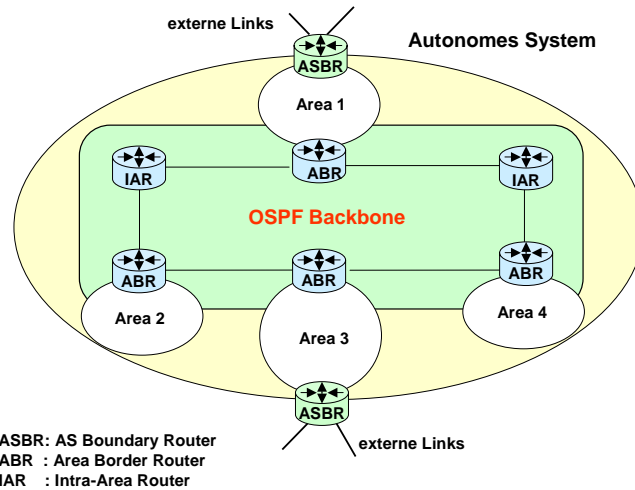
- jeder Übertragungsabschnitt im Netz entspricht einem Eintrag in der Datenbasis
- jeder Knoten kann kürzesten Pfad zu allen anderen Knoten im Netz berechnen
- Link State Advertisements (LSAs) werden im gesamten AS geflutet

Flooding-Algorithmus

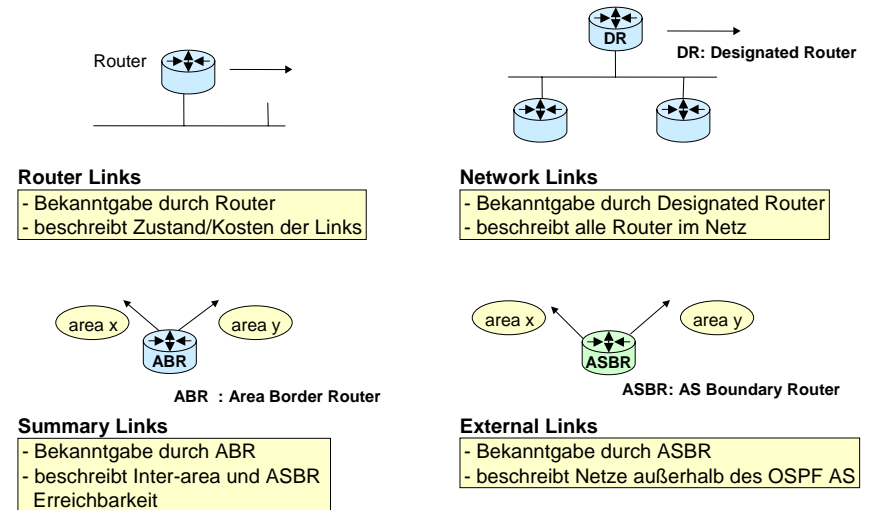
- Beim Empfangen einer LSA-Meldung wird Eintrag in der Datenbasis überprüft:
- falls Eintrag noch nicht vorhanden, dann Hinzufügen und Broadcasten
 - falls Eintrag vorhanden und neuer Wert niedriger als alter, dann Überschreiben und Broadcasten
 - falls Eintrag vorhanden und neuer Wert höher als alter, dann Übertragen des bereits gespeicherten Wertes über Eingangslink
 - falls beide Werte gleich sind, wird nichts unternommen

Wichtig: Updates müssen abgesichert werden
Hierzu verfügt OSPF (RFC 2178) über Quittungen pro Übertragungsabschnitt, Zeitgeber und Prüfsumme

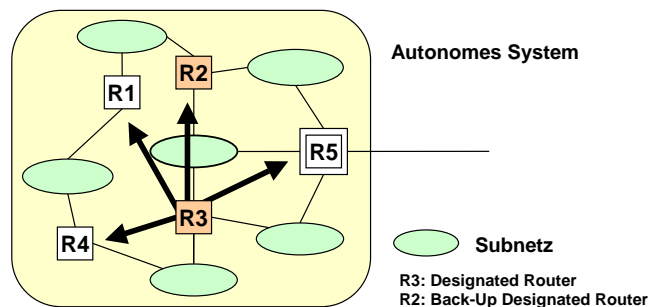
OSPF Netzstruktur



OSPF Link State Advertisements



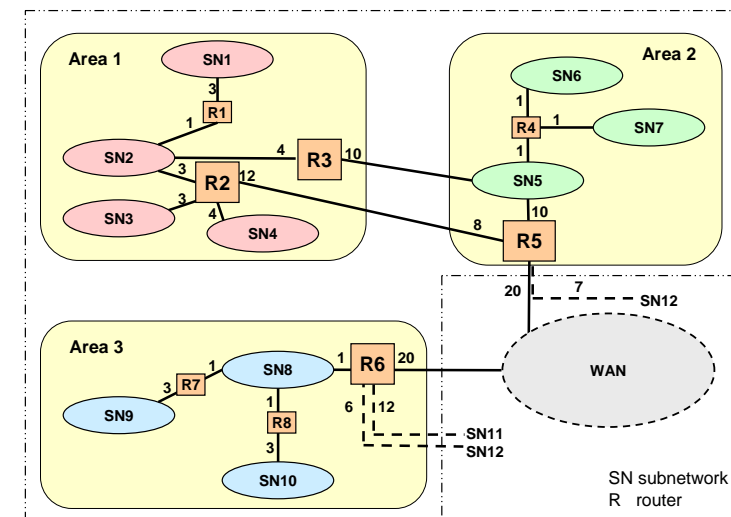
Routing Informationsaustausch



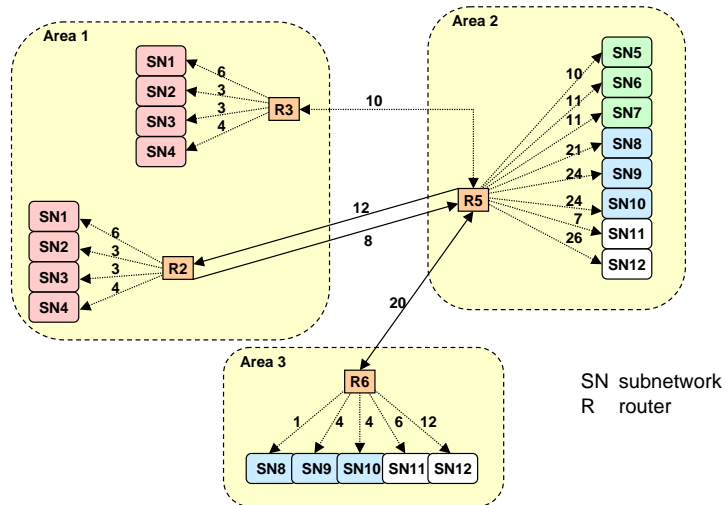
Anstatt Fluten an alle Router (Aufwand: $N \times (N-1)$ Meldungen):

- Bestimmung eines Designated Router (DR) und Back-Up Designated Router
- Jeder Router sendet die LSA-Meldung an ein Designated Router (DR)
- Designated Router leitet die LSA-Meldung an alle andere Router weiter
- Aufwand: $2N + (N-1) = (3N - 1)$ Meldungen

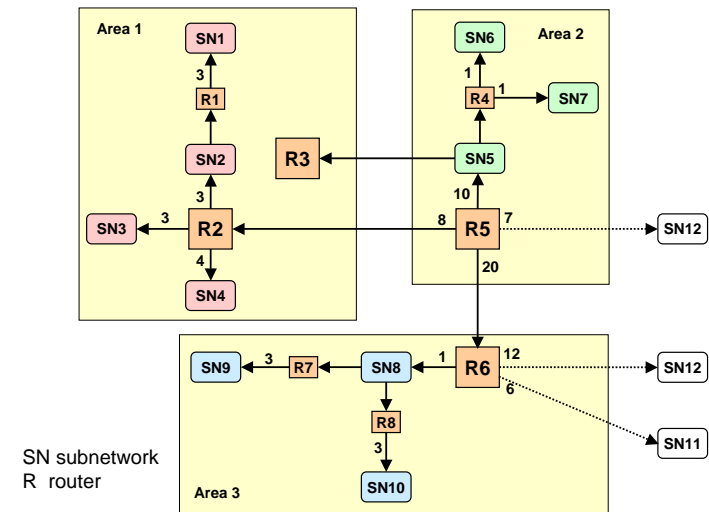
Autonomes System



Topologie-Datenbasis



Shortest-Path-First Baum für Router R5



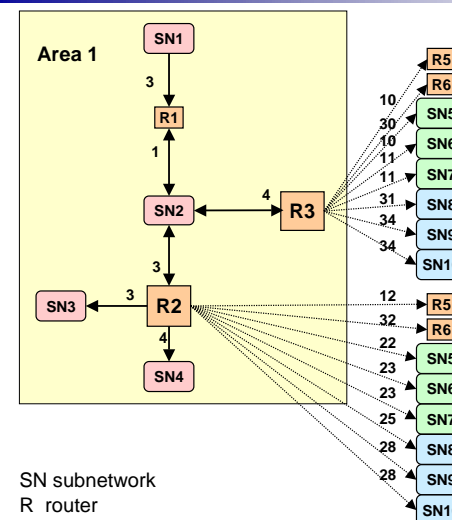
Topologie Datenbasis in Router R5

Ziel in AS	Nächste Router	Kosten
SN1	R2	14
SN2	R2	11
SN3	R2	11
SN4	R2	12
SN5	--	10
SN6	R4	11
SN7	R4	11
SN8	R6	21
SN9	R6	24
SN10	R6	24
R6	--	20

Ziel nicht in AS	Nächste Router	Kosten
SN11	R6	26
SN12	--	7

AS Autonomous System

Datenbasis für Area 1



Interior

Ziel in Area 1	Kosten über R2	Kosten über R3
SN1	6	7
SN2	3	4
SN3	3	7
SN4	4	8

Exterior

Ziel nicht in Area 1	Kosten über R2	Kosten über R3
R5	10	12
R6	30	32
SN5	10	22
SN6	11	23
SN7	11	23
SN8	31	25
SN9	34	28
SN10	34	28

Fortgeschrittene Funktionen in OSPF

Sicherheit

- Authentifizierung für Datenaustausch zwischen OSPF-Routern

Mehrere Pfade mit gleichen Kosten

- Verkehr kann bei gleichen Kosten über mehrere Pfade verteilt werden

Unterschiedliche Kostenmetriken für verschiedene Type-of-Service

- OSPF hat hierfür unterschiedliche Topologien und kann unterschiedliche Routen anbieten

Unterstützung von Unicast- und Multicasting

- Verwendet gleiche Topologie-Datenbasis

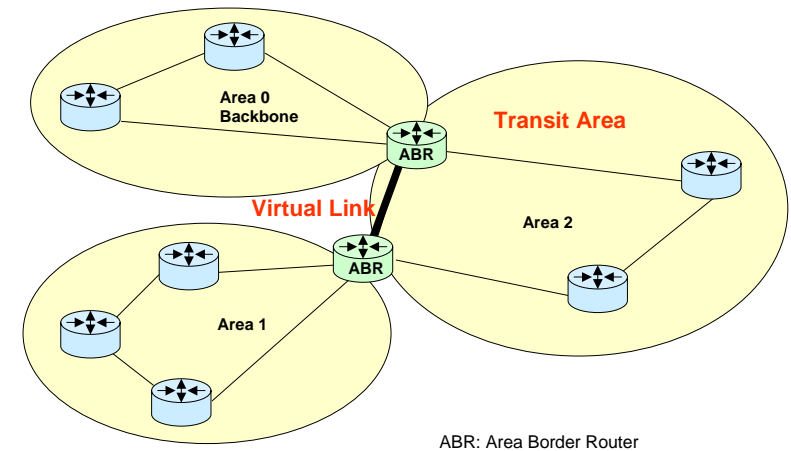
Hierarchie innerhalb eines AS

- AS kann in Areas untergliedert werden

Zwei Ebenen der Hierarchie

- Area
- Backbone (der die Areas verbindet)
 - Inter-Area-Kommunikation verläuft immer über Backbone-Area
 - Gehört ein Router zu zwei Areas, ist er automatisch ein Backbone-Router

Virtual Link / Transit Area



IS-IS

Ursprünglich im Rahmen von ISO/OSI für das dortige verbindungslose Netzprotokoll (CLNP: Connectionless Network Protocol) entworfen

Verwendung für IP möglich, aber nicht maßgeschneidert wie OSPF
Heute wird IS-IS noch von großen Providern eingesetzt

Eigenschaften

- Link-State-Protokoll
- Funktionalität ähnlich wie OSPF

Vorteile, die Befürworter nennen:

- Bessere Handhabung in sehr großen Areas
- Robusteres Protokoll für das Fluten
- Teilweise qualitativ bessere Implementierung in manchen Routern

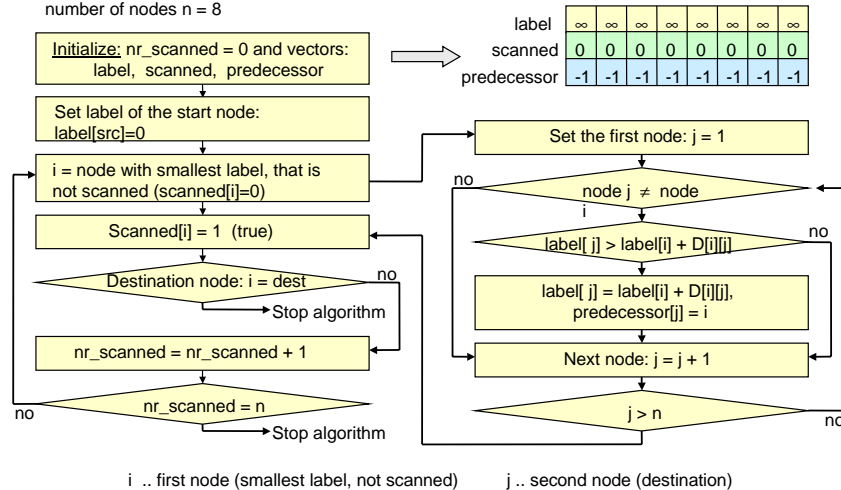
Derzeit: Weiterentwicklung und Verbesserungen für den Einsatz mit IP in der IS-IS Working Group der IETF

Dijkstra's Shortest Path Algorithmus

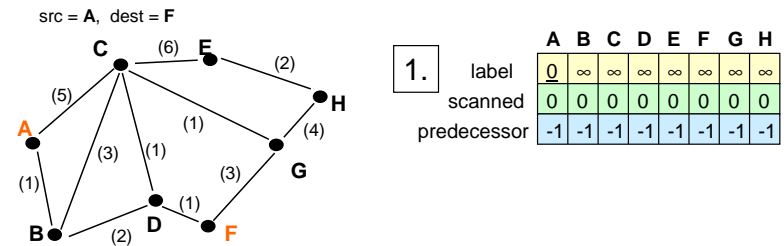
- Realisierung durch Forward Search Algorithmus
- Jeder Router hält 2 Listen:
Confirmed und Tentative mit Einträgen der Form {Ziel, Kosten, nächster Knoten}
- Algorithmus
 1. Initialisierung von Confirmed mit eigenem Eintrag (Kosten 0)
 2. Selektiere den zuletzt in Confirmed aufgenommenen (Next)
 3. Berechne die Kosten zu allen Nachbarn von Next:
 $\text{Kosten} = \text{Kosten zu Next} + \text{Kosten von Next zu Nachbar}$
 - a. Falls Nachbar nicht in Confirmed:
Aufnahme des Nachbarn in Tentative
 - b. Falls Nachbar in Tentative: Ersetzen des Eintrags, falls die Kosten niedriger als die des Eintrags sind.
 4. Stop, falls Tentative leer ist, sonst: übernehme in Confirmed den Eintrag mit den geringsten Kosten aus Tentative, gehe zu 2

Shortest Path Algorithm (Dijkstra)

source: src, destination: dest, distance matrix: D,
number of nodes n = 8

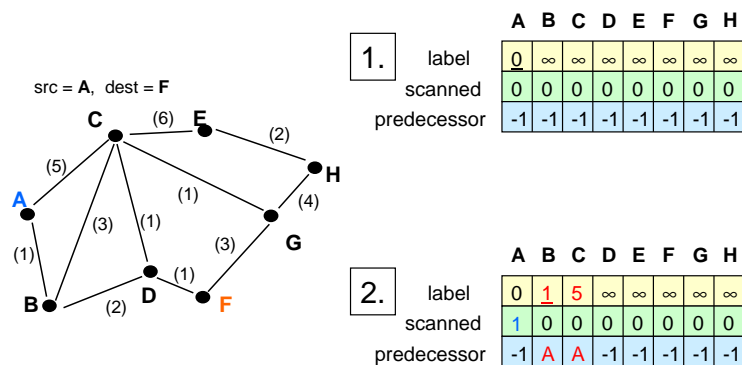


Shortest Path (1)

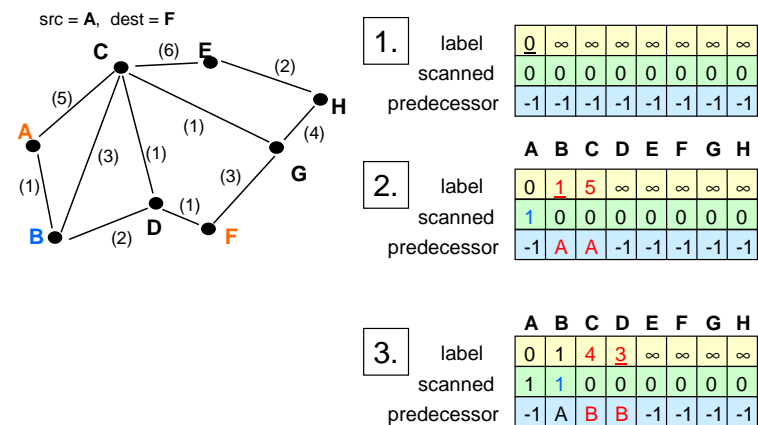


smallest not scanned label (0) is **scanned next**
predecessor and label are replaced if the new label is smaller

Shortest Path (2)

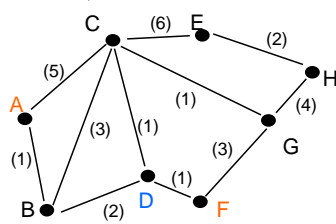


Shortest Path (3)



Shortest Path (4)

src = A, dest = F



1. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	∞	∞	∞	∞	∞	∞	∞
scanned	0	0	0	0	0	0	0	0
predecessor	-1	-1	-1	-1	-1	-1	-1	-1

2. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	5	∞	∞	∞	∞	∞
scanned	1	0	0	0	0	0	0	0
predecessor	-1	A	A	-1	-1	-1	-1	-1

4. label scanned predecessor

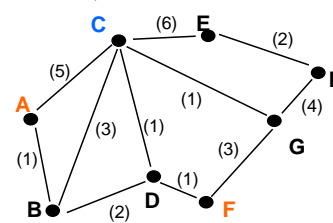
	A	B	C	D	E	F	G	H
label	0	1	4	3	∞	4	∞	∞
scanned	1	1	0	1	0	0	0	0
predecessor	-1	A	B	B	-1	D	-1	-1

3. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	∞	∞	∞	∞
scanned	1	1	0	0	0	0	0	0
predecessor	-1	A	B	B	-1	-1	-1	-1

Shortest Path (5)

src = A, dest = F



5. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	∞
scanned	1	1	1	1	0	0	0	0
predecessor	-1	A	B	B	C	D	C	-1

F has the smallest not scanned label!

1. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	∞	∞	∞	∞	∞	∞	∞
scanned	0	0	0	0	0	0	0	0
predecessor	-1	-1	-1	-1	-1	-1	-1	-1

2. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	5	∞	∞	∞	∞	∞
scanned	1	0	0	0	0	0	0	0
predecessor	-1	A	A	-1	-1	-1	-1	-1

3. label scanned predecessor

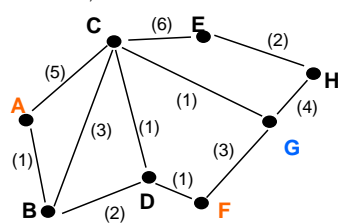
	A	B	C	D	E	F	G	H
label	0	1	4	3	∞	∞	∞	∞
scanned	1	1	0	0	0	0	0	0
predecessor	-1	A	B	B	-1	-1	-1	-1

4. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	∞	4	∞	∞
scanned	1	1	0	1	0	0	0	0
predecessor	-1	A	B	B	-1	D	-1	-1

Shortest Path (6)

src = A, dest = F



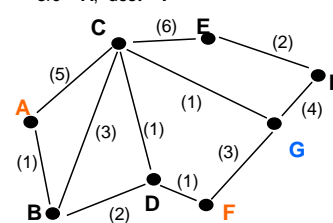
6. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	∞
scanned	1	1	1	1	0	1	0	0
predecessor	-1	A	B	B	C	D	C	-1

Complete table for source A

Shortest Path (7)

src = A, dest = F



6. label scanned predecessor

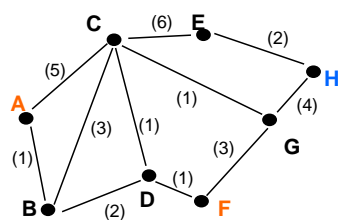
	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	∞
scanned	1	1	1	1	0	1	0	0
predecessor	-1	A	B	B	C	D	C	-1

7. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	9
scanned	1	1	1	1	0	1	1	0
predecessor	-1	A	B	B	C	D	C	G

Shortest Path (8)

src = A, dest = F



6. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	∞
scanned	1	1	1	1	0	1	0	0
predecessor	-1	A	B	B	C	D	C	-1

7. label scanned predecessor

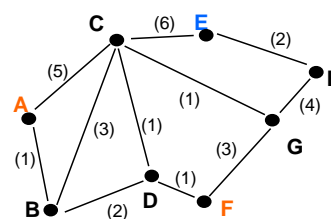
	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	9
scanned	1	1	1	1	0	1	1	0
predecessor	-1	A	B	B	C	D	C	G

8. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	9
scanned	1	1	1	1	0	1	1	1
predecessor	-1	A	B	B	C	D	C	G

Shortest Path (9)

src = A, dest = F



6. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	∞
scanned	1	1	1	1	0	1	0	0
predecessor	-1	A	B	B	C	D	C	-1

7. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	9
scanned	1	1	1	1	0	1	1	0
predecessor	-1	A	B	B	C	D	C	G

9. label scanned predecessor

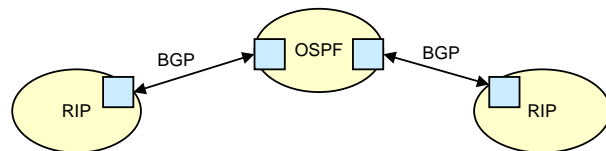
	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	9
scanned	1	1	1	1	1	1	1	1
predecessor	-1	A	B	B	C	D	C	G

8. label scanned predecessor

	A	B	C	D	E	F	G	H
label	0	1	4	3	10	4	5	9
scanned	1	1	1	1	0	1	1	1
predecessor	-1	A	B	B	C	D	C	G

Shortest path: FDBA → ABDF

Border Gateway Protocol (BGP)



- OSPF und RIP nur innerhalb der einzelnen Domänen (begrenzte Skalierbarkeit !)
- BGP als separates Routing-Protokoll zwischen Domänen (= autonome Systeme, AS)
 - Inter-Domain-Routing-Protokoll
 - Distanz-Vektor-Prinzip
 - Skalierbarkeit durch Hierarchiebildung
 - Berücksichtigung von administrativen Beschränkungen (Policies)

Inter-AS-Routing: BGP

Pfad-Vektor-Protokoll

- Erweiterung zum Distanz-Vektor
- BGP verbreitet keine Metriken wie Kosten etc. sondern Pfade
- Pfade garantieren Schleifenfreiheit
- Policies sind hier ausschlaggebend für die Wegewahl

Kommunikation zwischen BGP-Instanzen (gesichert durch TCP)

OPEN

- Aufbau einer Verbindung zum Kommunikationspartner Authentisierung

UPDATE

- Bekanntgabe eines neuen Pfads oder Zurücknahme eines alten Pfads

KEEP ALIVE

- Hält Verbindung aufrecht in Abwesenheit von UPDATE-Dateneinheiten Quittung zu einem OPEN-Request

NOTIFICATION

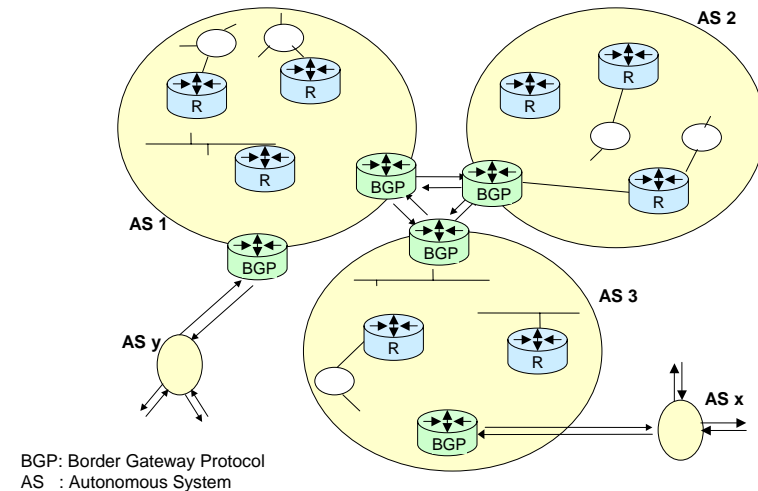
- Mitteilung von Fehlern in vorangegangenen Dateneinheiten Abbau einer Verbindung

BGP: Border Gateway Protocol

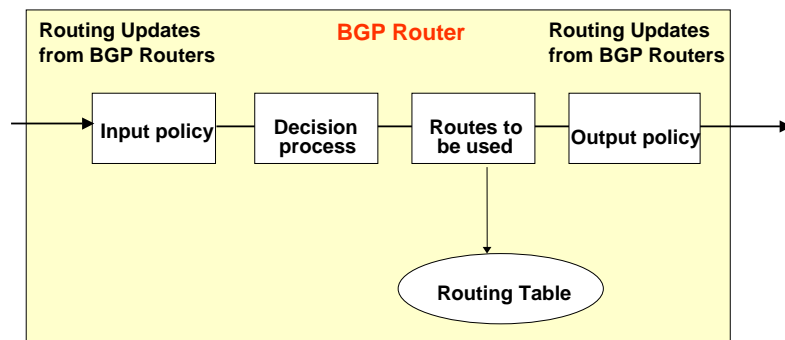
- Router and AS relationship
- Message flow
- Message header
- Open message
- Update message
- Exchanging NLRI
- Path attributes
- Exchanging withdrawn routes
- Notification message

NLRI: Network Layer Reachability Information

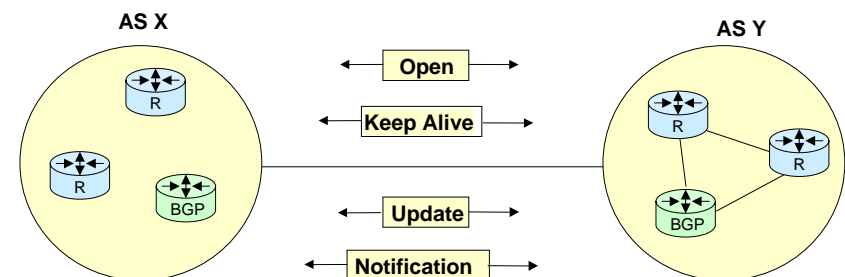
BGP Router and AS Relationship



BGP Process and Routing Policies



BGP Message Flow



- OPEN: Aufbau einer TCP-Verbindung zum Austausch von Routing-Informationen zwischen 2 BGP-Routern
- UPDATE: enthält Pfad-Attribute und Liste erreichbarer Netze
- NOTIFICATION: Mitteilung von Fehlern
- KEEPALIVE: Erreichbarkeitskontrolle der Nachbar-BGP-Router

