Quelle: Computernetzwerke: Der Top-Down-Ansatz, 6. Ausgabe
Jim Kurose, Keith Ross; Pearson, März 2014

Technische Grundlagen der Informatik 2 – Teil 6: Netzwerkschicht (Layer 3)

Philipp Rettberg / Sebastian Harnau

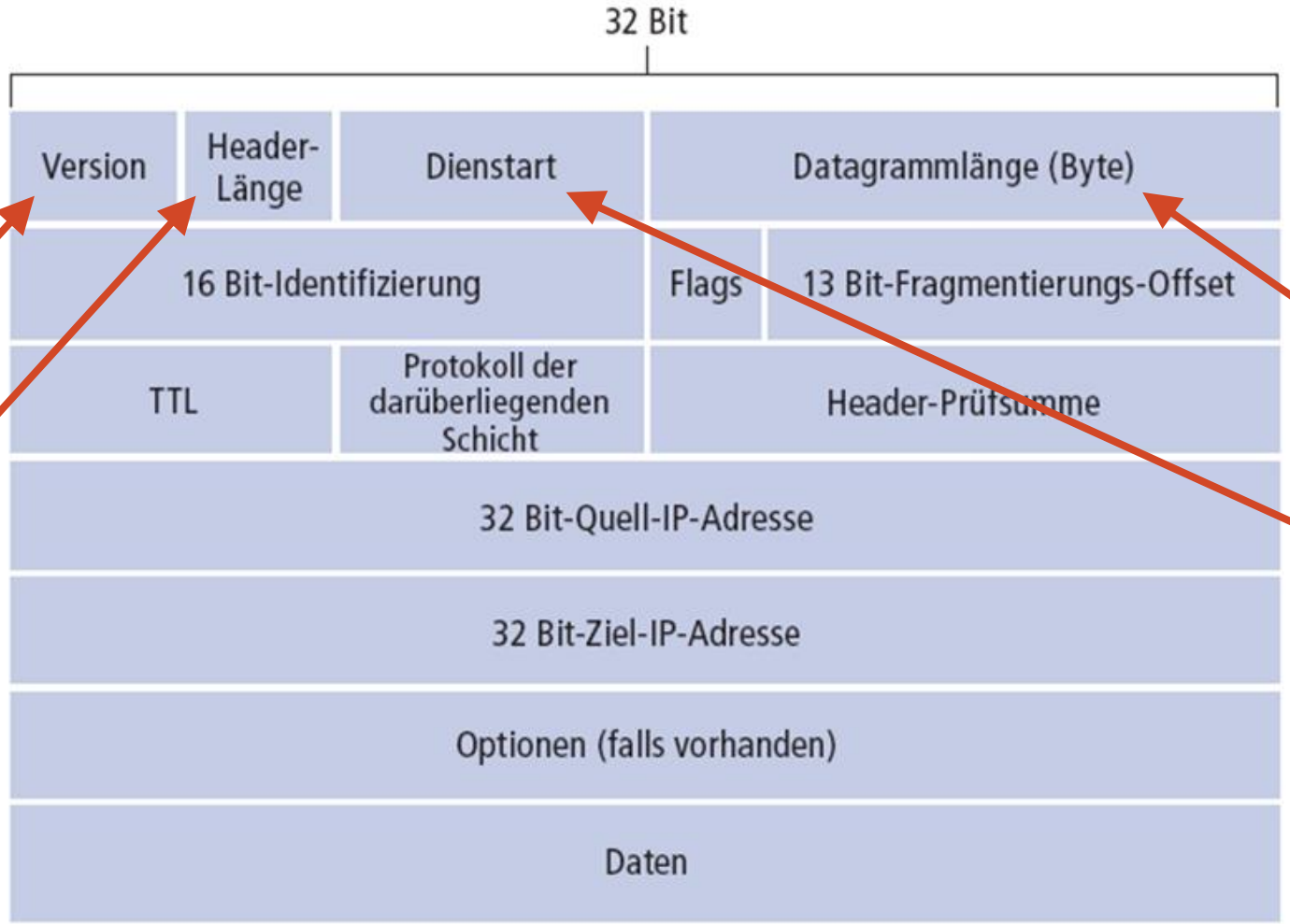
Block 11/18

Netzwerkschicht (Layer 3)

IP, Router, Routing…

IPv4 Datagrammformat

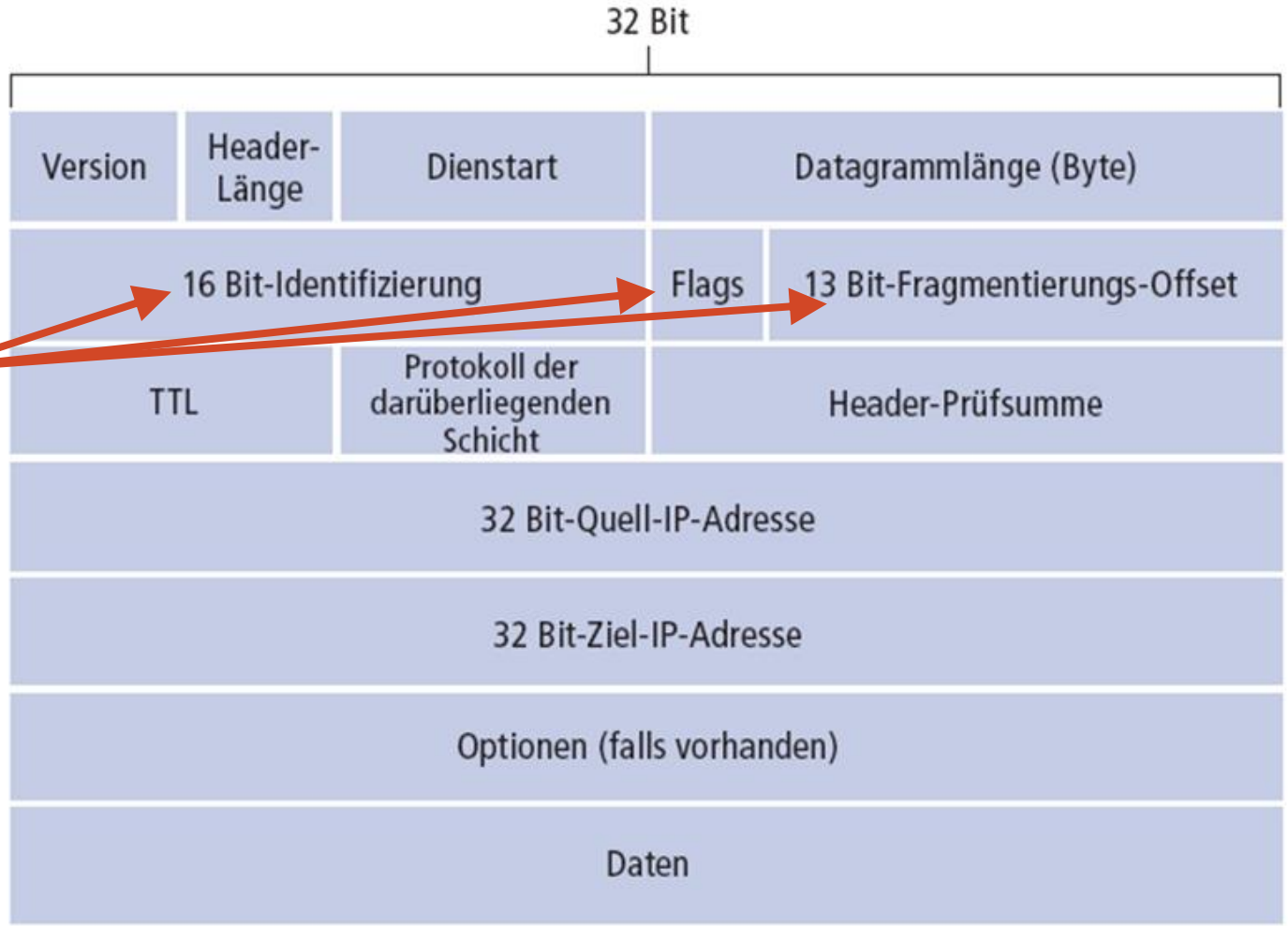
- IP-Protokoll-version
- Header-Länge in Byte



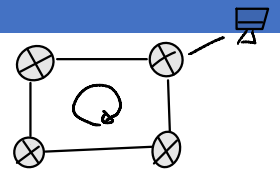
- Größe des Datagramms in Byte
- Hinweise zur Behandlung der Daten (meist ignoriert)

IPv4 Datagrammformat

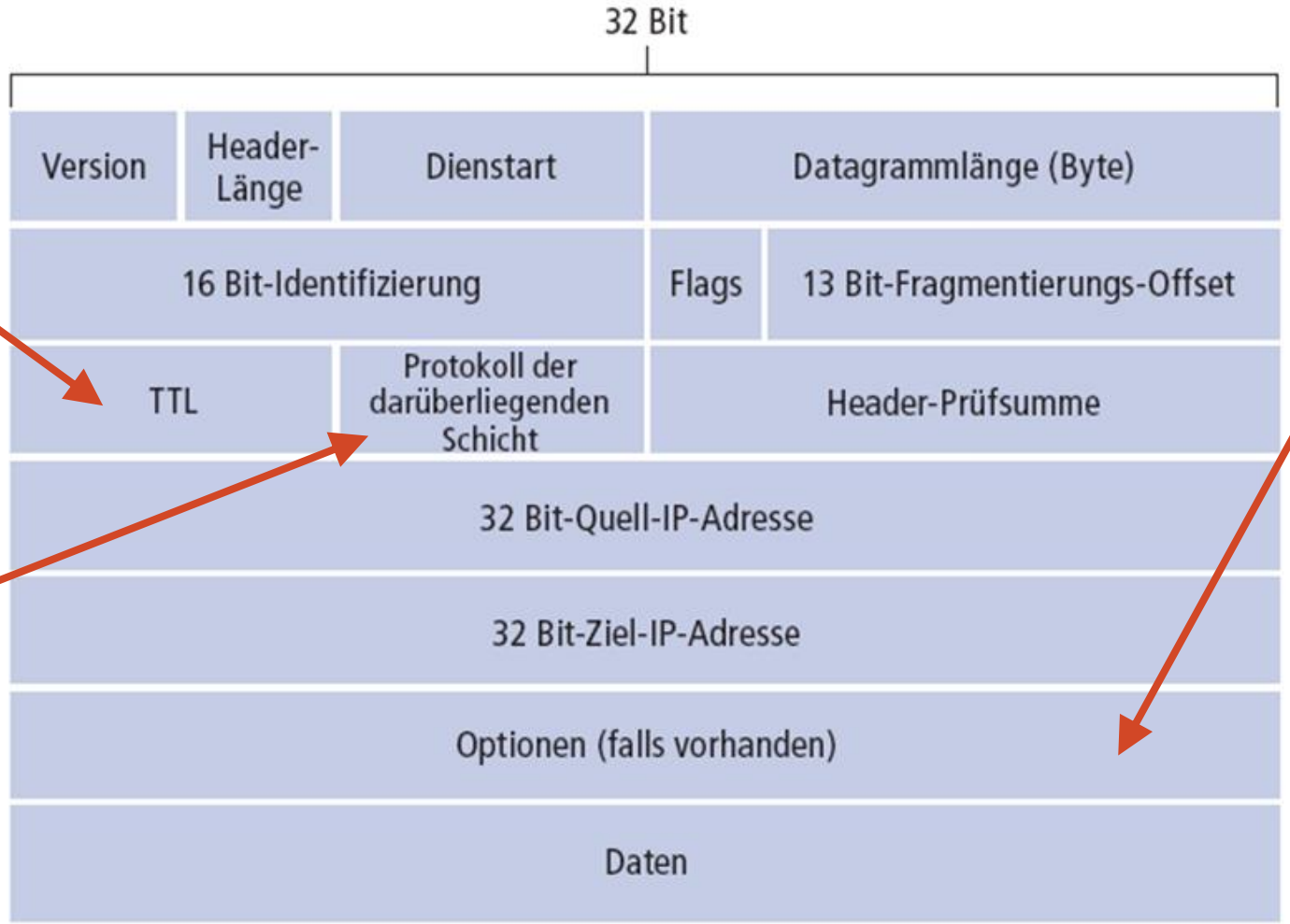
- Für die Fragmentierung von Datagrammen



IPv4 Datagrammformat



- Maximale Anzahl der noch zu durch-laufenden Router (wird von jedem Router dekrementiert)
- Protokoll der nächsthöheren Schicht, an die das Datagramm ausgeliefert werden soll

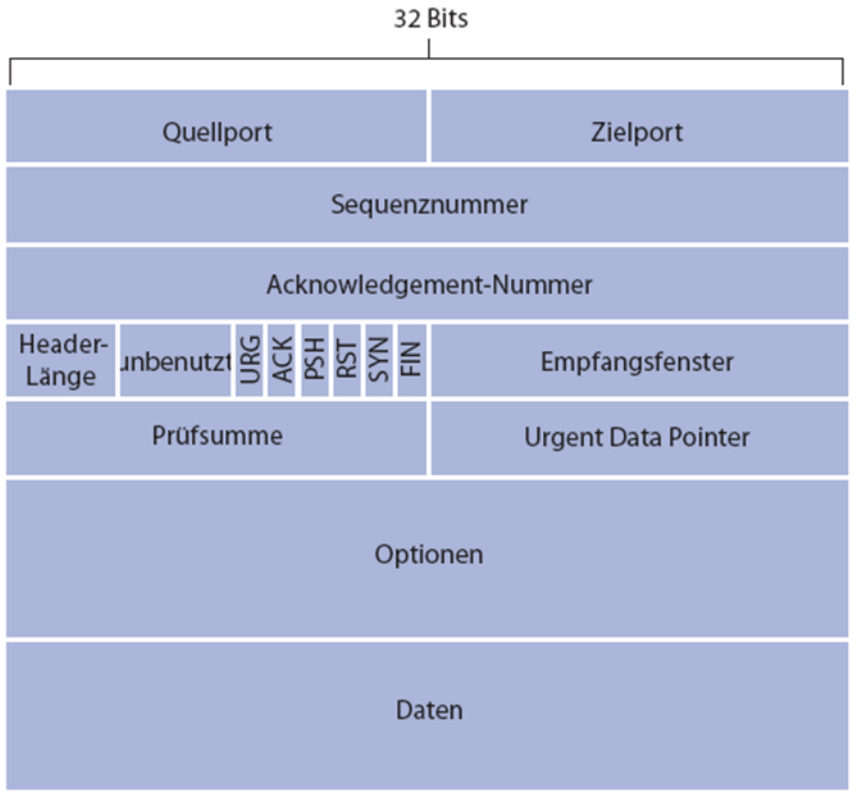


Beispiele:

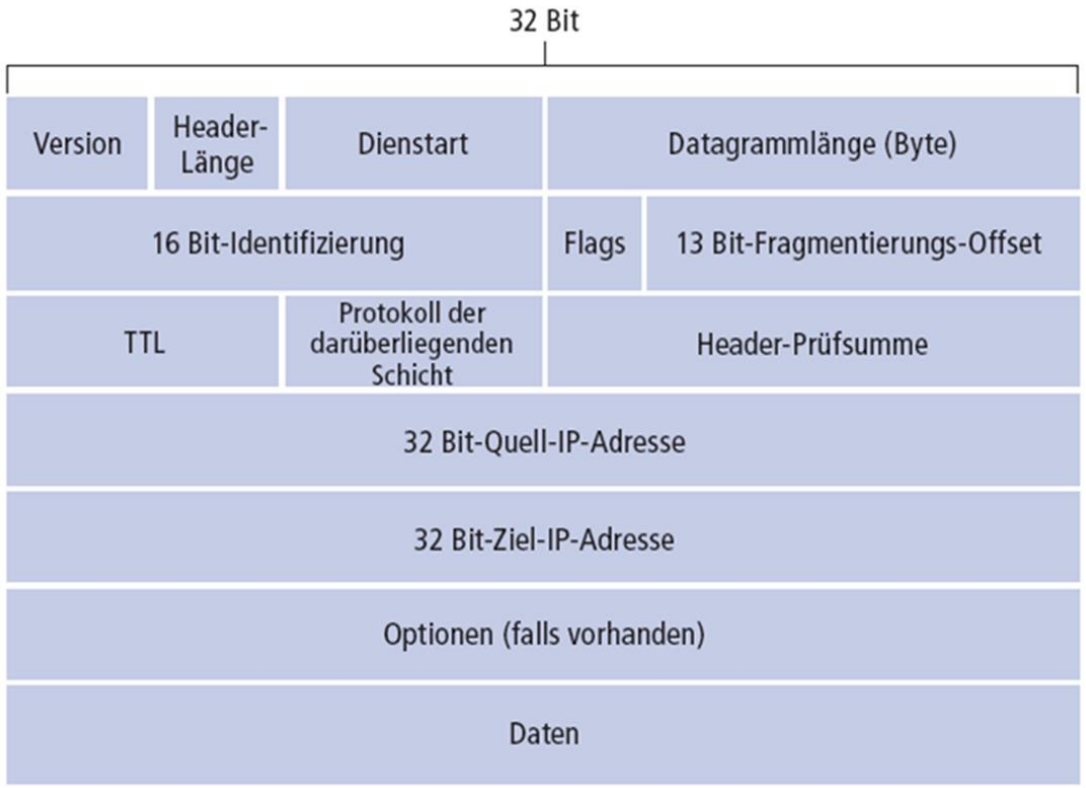
- Zeitstempel;
- Aufzeichnen der durchlaufenen Router
- Router, die durchlaufen werden sollen

Header TCP / IP

TCP



IP



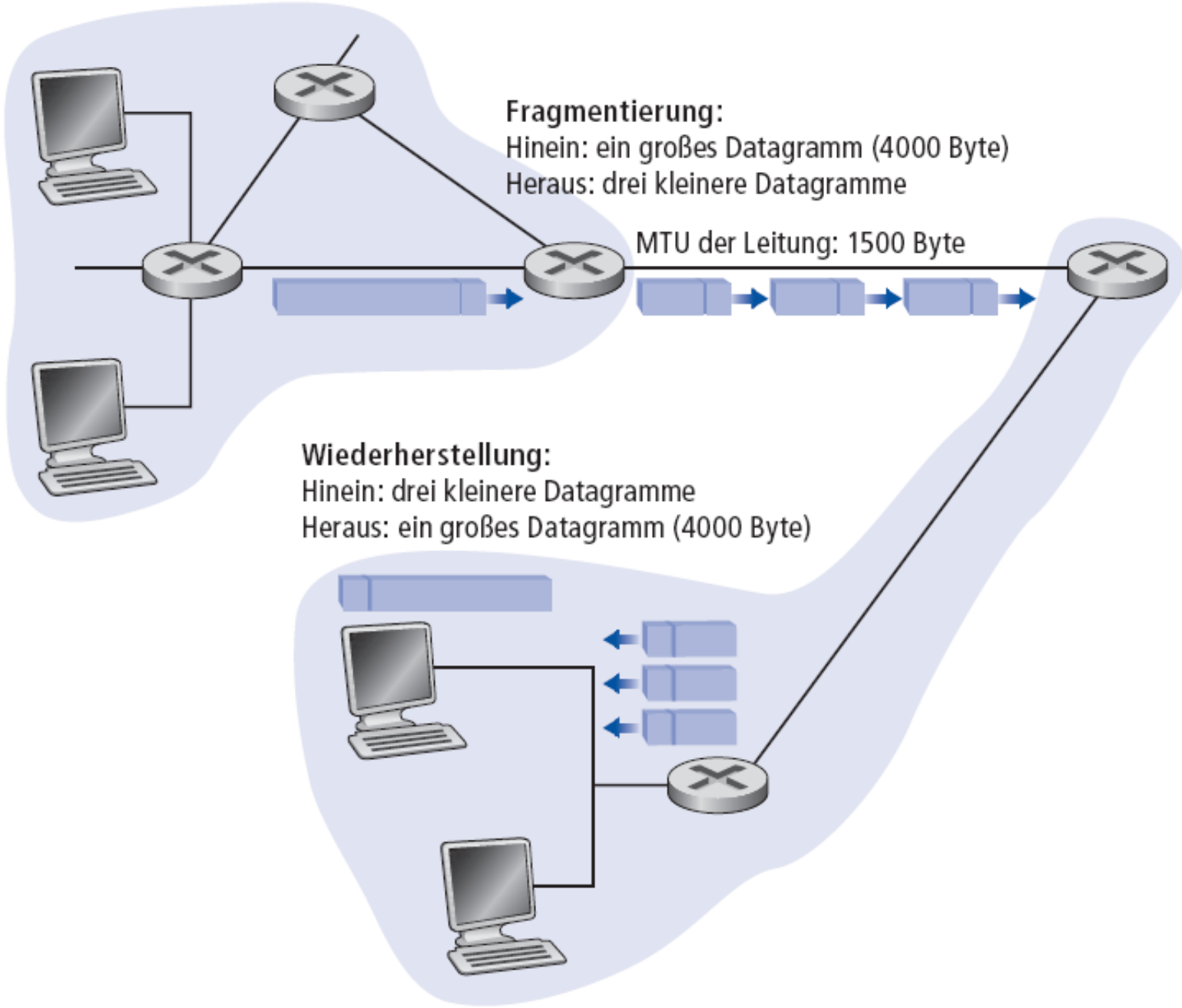
Wieviel Overhead entsteht (mindestens) bei der Nutzung von TCP/IP?

IP-Fragmentierung

- Links haben eine Maximalgröße für Frames
- Diese nennt man **Maximum Transmission Unit** (MTU)
- Verschiedene Links haben unterschiedliche MTUs
- IP-Datagramme müssen unter Umständen aufgeteilt werden

IP-Fragmentierung

- Aufteilung (Fragmentierung) erfolgt in den Routern
- Zusammensetzen (Reassembly) erfolgt beim Empfänger
- IP-Header enthält die notwendigen Informationen hierzu



IP-Fragmentierung

Beispiel:

IP-Datagramm mit 4000 Byte (inklusive 20 Byte IP-Header)

MTU des nächsten Links = 1500 Byte

Fragment	Bytes	ID	Offset	Flag
1. Fragment	1.480 Byte im Daten-feld des IP-Datagramms	Identifizierung = 777	Offset = 0 (d.h., die Daten sollten beginnend bei Byte 0 eingefügt werden)	Flag = 1 (d.h., da kommt noch mehr)
2. Fragment	1.480 Datenbytes	Identifizierung = 777	Offset = 185 (d.h., die Daten sollten bei Byte 1.480 beginnend einge-fügt werden; beachten Sie, dass 185 · 8 = 1.480)	Flag = 1 (d.h., da kommt noch mehr)
3. Fragment	1.020 Datenbytes (= 3.980 – 1.480 – 1.480)	Identifizierung = 777	Offset = 370 (d.h., die Daten sollten beginnend bei Byte 2.960 eingefügt werden; beachten Sie, dass 370 · 8 = 2.960)	Flag = 0 (d.h., es ist das letzte Fragment)

IP-Fragmentierung

Praktisch:

- Endsystem/Anwendung muss sich keine Gedanken über die Größe von MTUs verschiedener Links auf dem Weg vom Sender zum Empfänger machen
- Entspricht dem Prinzip einer geschichteten Architektur

Aber:

- Aufwand in den Routern
- Wenn ein Fragment verloren geht, ist das ganze Datagramm verloren
- Daher: **Fragmentation considered harmful!**

Vermeidung von Fragmentierung

Lösung: Bestimmen der kleinsten MTU des Weges (Path MTU)

- Setze DF (Don't-Fragment-Bit) im Header des IP-Paketes
- Wenn fragmentiert werden soll, wird das Paket verworfen und der Sender per ICMP benachrichtigt
- Sender wählt dann kleinere MTU
- Wiederholen, bis akzeptable MTU gefunden wurde

ICMP: Internet Control Message Protocol

- Wird von Hosts und Routern verwendet, um Informationen über das Netzwerk selbst zu verbreiten
  - Fehlermeldungen: Host, Netzwerk, Port, Protokoll nicht erreichbar
  - Echo-Anforderung und Antwort (von ping genutzt)
- Gehört zur Netzwerkschicht, wird aber in IP-Datagrammen transportiert
- ICMP-Nachricht: Type, Code und die ersten 8 Byte des IP-Datagramms, welches die Nachricht ausgelöst hat

Type	Code	Bescheibung
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

Traceroute – Infos über Router auf dem Weg zum Ziel

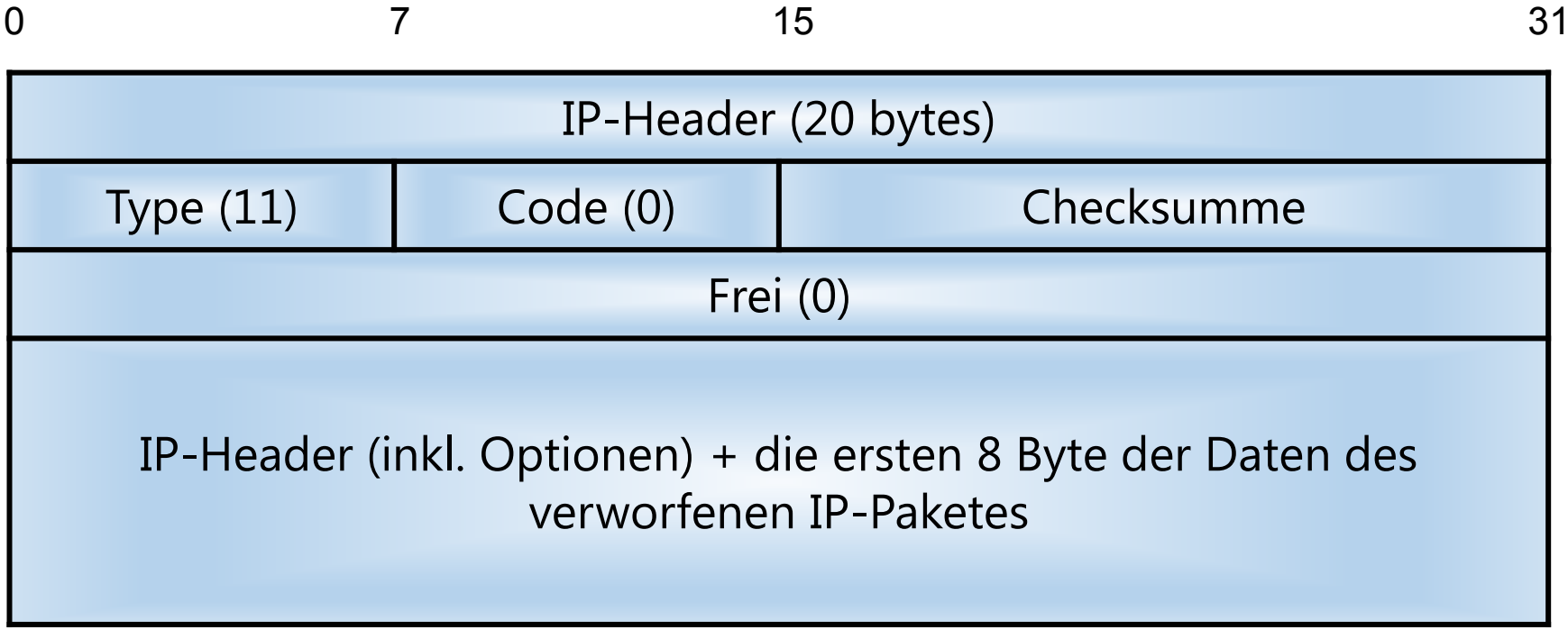
Aufgabe:

- Traceroute bestimmt Informationen über alle Router, die auf dem Weg zu einer IP-Adresse liegen
- Dabei wird auch die Round-Trip-Zeit zu jedem Router bestimmt

Funktionsweise:

- Traceroute schickt ein UDP-Paket an die Adresse, für die der Weg untersucht werden soll; TTL im IP-Header wird auf 1 gesetzt
- Der erste Router verwirft das IP-Paket (TTL = 1!) und schickt eine ICMP-Time-Exceeded-Fehlermeldung an den Absender
- Traceroute wiederholt dies mit TTL = 2 etc.

ICMP Time-Exceeded-Nachricht



ICMP Traceroute

- Wie erkennt man, ob das Paket schließlich beim Empfänger angekommen ist?
- Traceroute sendet UDP-Pakete an einen Port, der **wahrscheinlich** nicht verwendet wird, und erwartet eine ICMP-Port-Unreachable-Nachricht vom Empfänger!
- Traceroute ist ein „Hack“!
- Demo:

```
> traceroute <host>
```

- Traceroute berichtet die IP-Adresse des Interface, auf dem das Paket ankommt!

Lab: Traceroute

Windows: tracert <dest host>

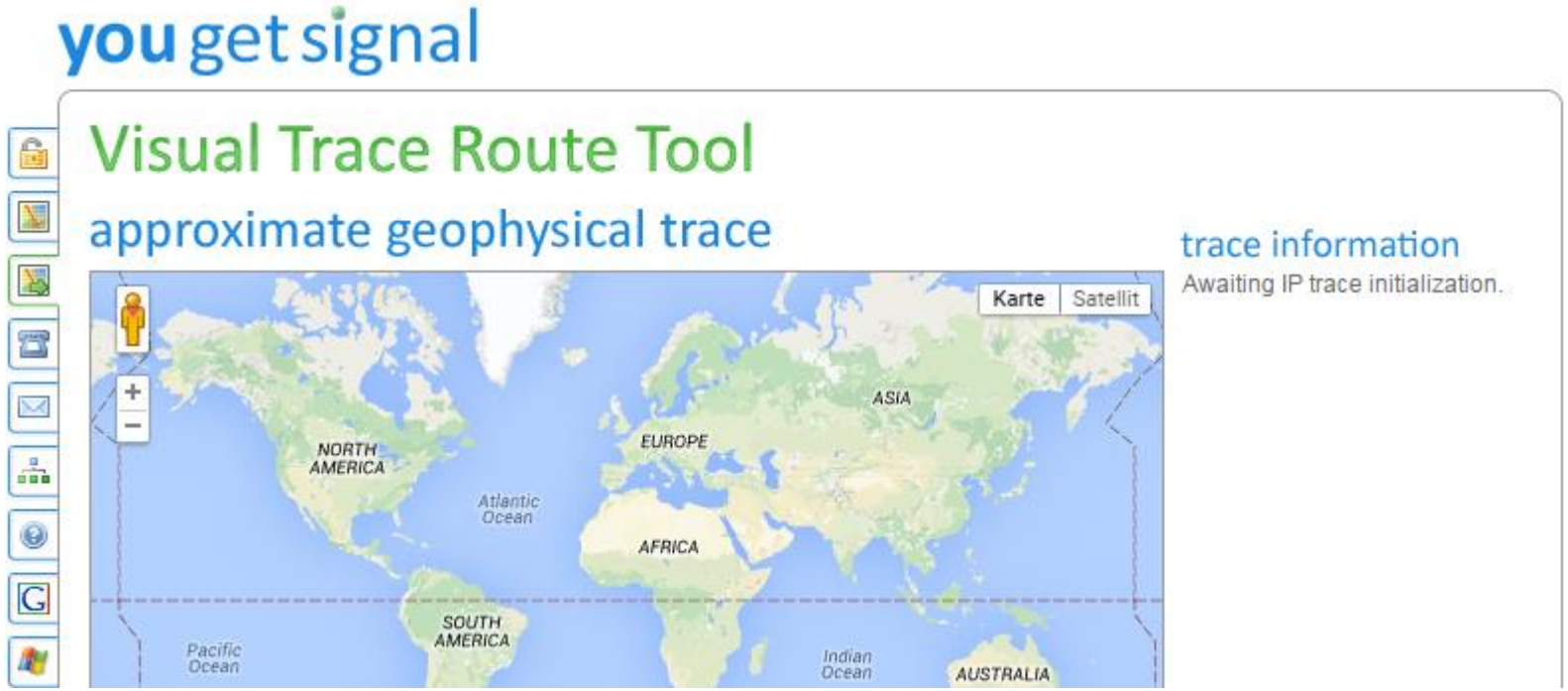
Linux/Mac: traceroute <dest host>

Destinations:

- apple.com
- google.com
- asustek.tw
- …

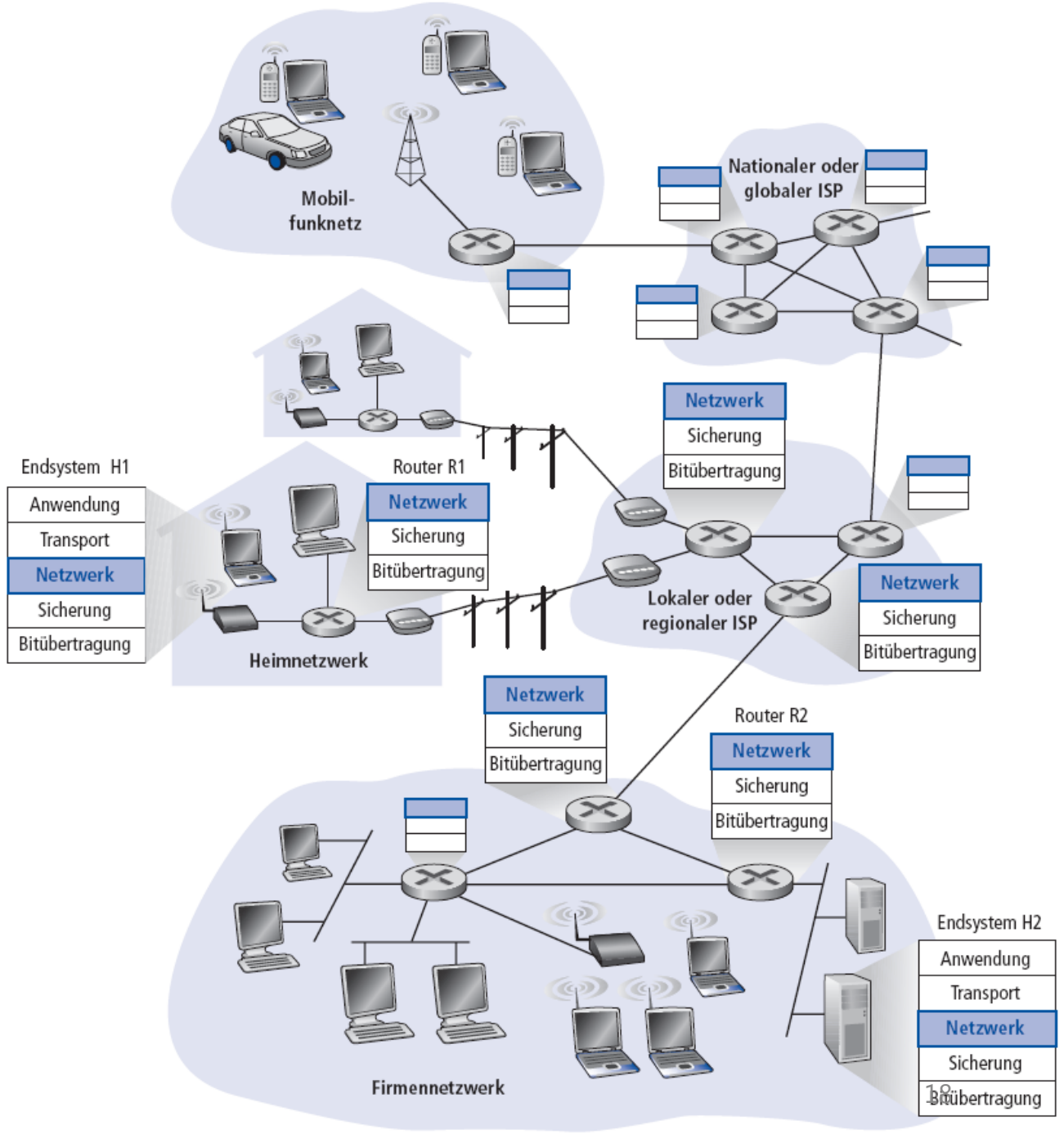
Lab: VisualTraceRoute

http://www.yougetsignal.com/tools/visual-tracert/



Netzwerkschicht

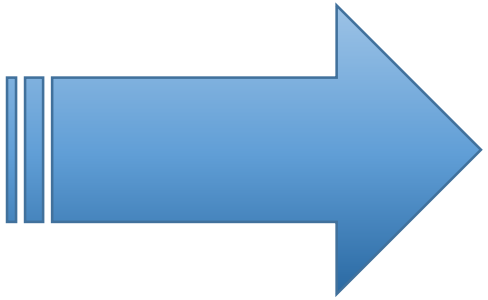
- **Auch:** Vermittlungsschicht oder Network Layer
- Daten von der nächsthöheren Schicht (Transportschicht) des Senders entgegennehmen
- In Datagramme verpacken
- Durch das Netzwerk leiten
- Auspacken des Vermittlungspakets beim Empfänger
- Ausliefern der Daten an die nächsthöhere Schicht (Transportschicht) des Empfängers
- Netzwerkschicht existiert in jedem Host und Router!



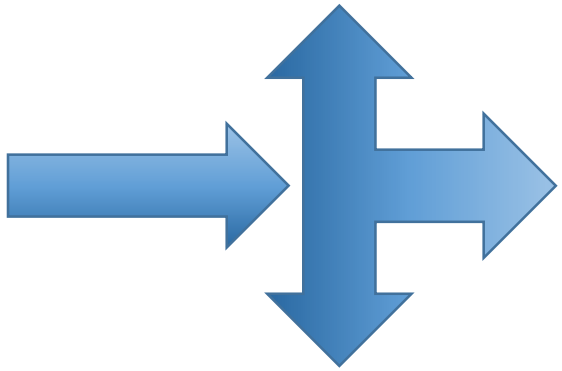
Funktionen der Netzwerkschicht

Die Netzwerkschicht kümmert sich im Wesentlichen um zwei zentrale Funktionen:

Weiterleiten



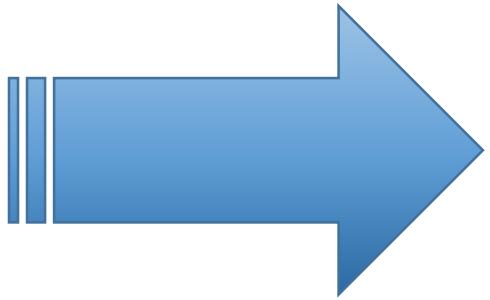
Wegewahl



Funktionen der Netzwerkschicht: Forwarding

Weiterleiten von Paketen (Forwarding):

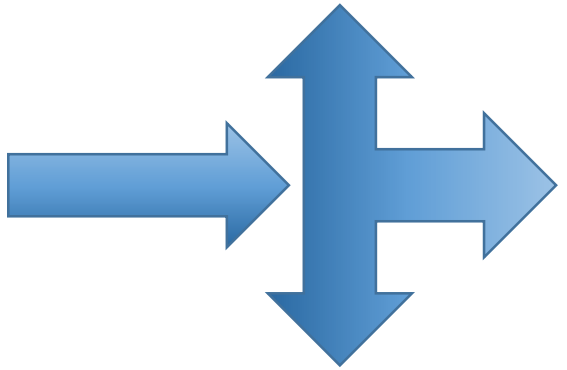
- Router nimmt Paket auf einer Eingangsleitung entgegen
- Router bestimmt die Ausgangsleitung anhand lokaler Informationen (z.B. Routing-Tabelle)
- Router legt das Paket auf die Ausgangsleitung



Funktionen der Netzwerkschicht: Routing

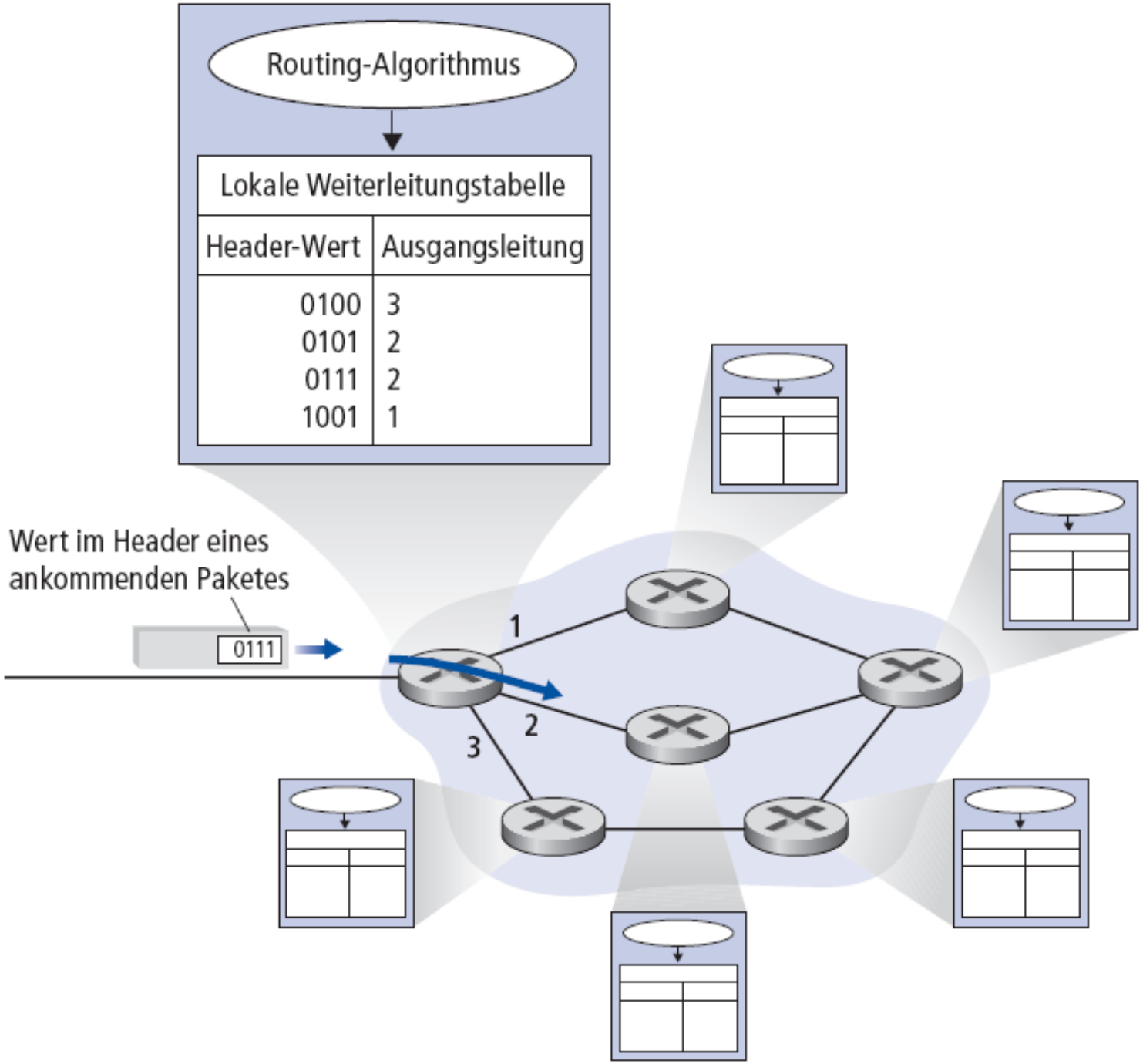
Wegewahl (Routing):

- Router kommunizieren miteinander, um geeignete Wege durch das Netzwerk zu bestimmen
- Als Ergebnis erhalten sie Informationen, wie welches System im Netzwerk zu erreichen ist (z.B. wird eine Routingtabelle mit Einträgen gefüllt)



Routing und Forwarding

Das Zusammenspiel von Routing und Forwarding lässt sich wie folgt skizzieren:



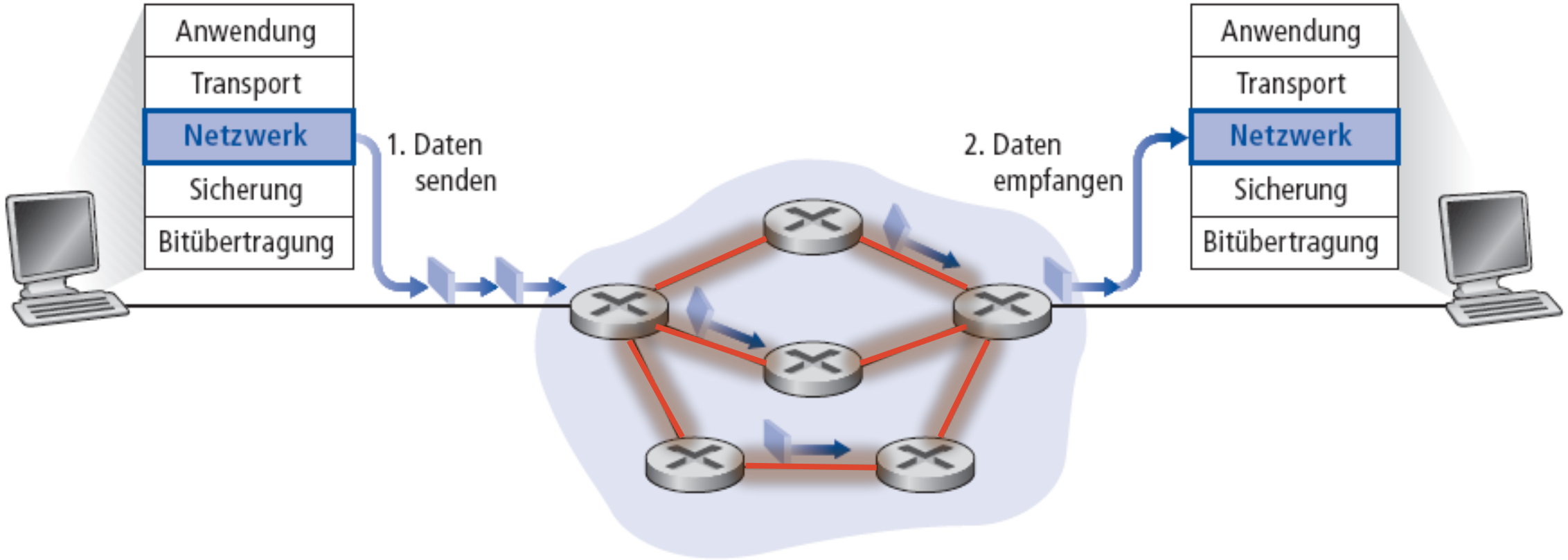
Datagrammnetzwerk

Dem Internet liegt ein **Datagrammnetzwerk** zugrunde. Ein Datagrammnetzwerk verwendet eine verbindungslose Netzwerkschicht. Dies bedeutet im Speziellen:

- Kein Verbindungsaufbau auf der Netzwerkschicht
- Router halten keinen Zustand für Ende-zu-Ende-Verbindungen
    - Auf Netzwerkebene gibt es das Konzept einer “Verbindung” nicht
- Pakete werden unter Verwendung einer Zieladresse weitergeleitet
    - Pakete für dasselbe Sender-Empfänger-Paar können unterschiedliche Pfade nehmen

Hinweis: Es gibt auch eine verbindungsorientierte Netzwerkschicht, die „virtuelle Leitungen“ bereitstellt (Stichwort z. B.: ATM). Aufgrund der geringen Relevanz soll hier nicht weiter darauf eingegangen werden.

Datagrammnetzwerk



Weiterleitungstabelle

Zieladressbereich	Schnittstelle
11001000 00010111 00010000 00000000	
bis	0
11001000 00010111 00010111 11111111	
11001000 00010111 00011000 00000000	
bis	1
11001000 00010111 00011000 11111111	
11001000 00010111 00011001 00000000	
bis	2
11001000 00010111 00011111 11111111	
sonst	3

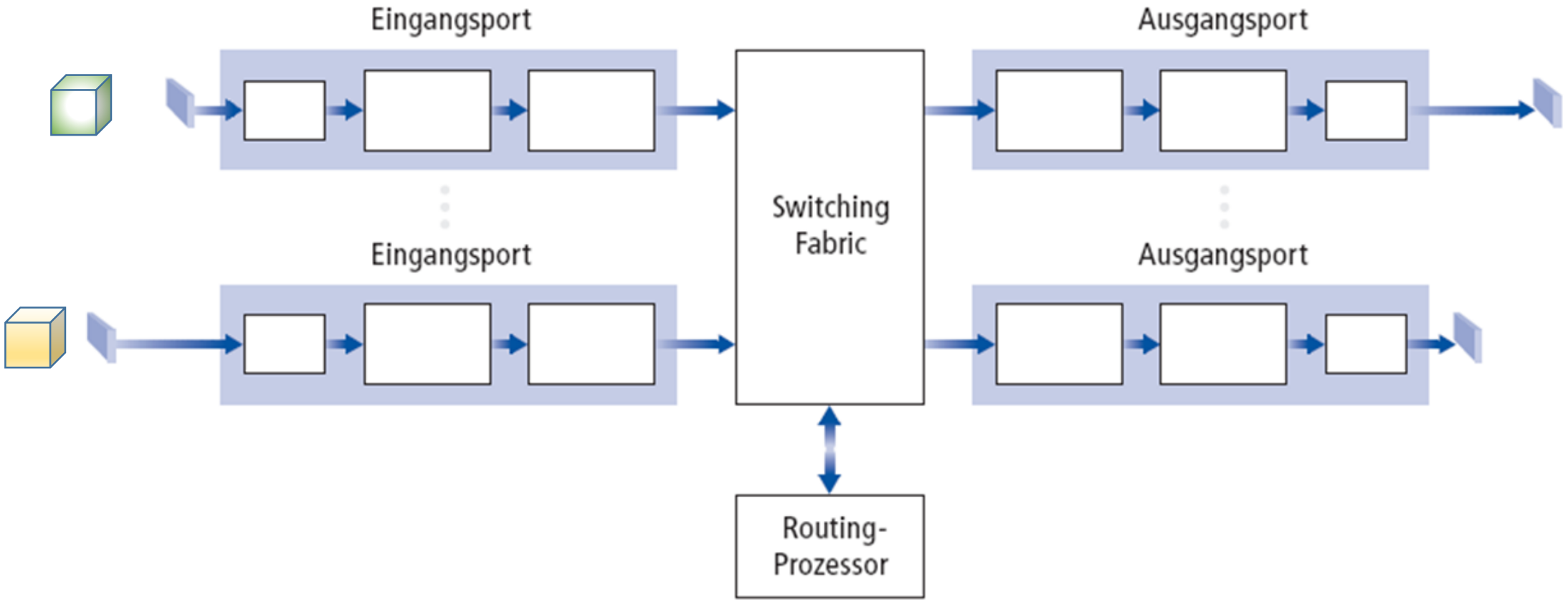
Longest Prefix Matching

Passender Präfix	Schnittstelle
11001000 00010111 00010	0
11001000 0 0010111 00011000	1
11001000 00010111 00011	2
sonst	3

Beispiele:

- Adresse: 11001000 00010111 00010110 10100001
- Adresse: 11001000 00010111 00011000 10101010

Übersicht: Routerarchitektur

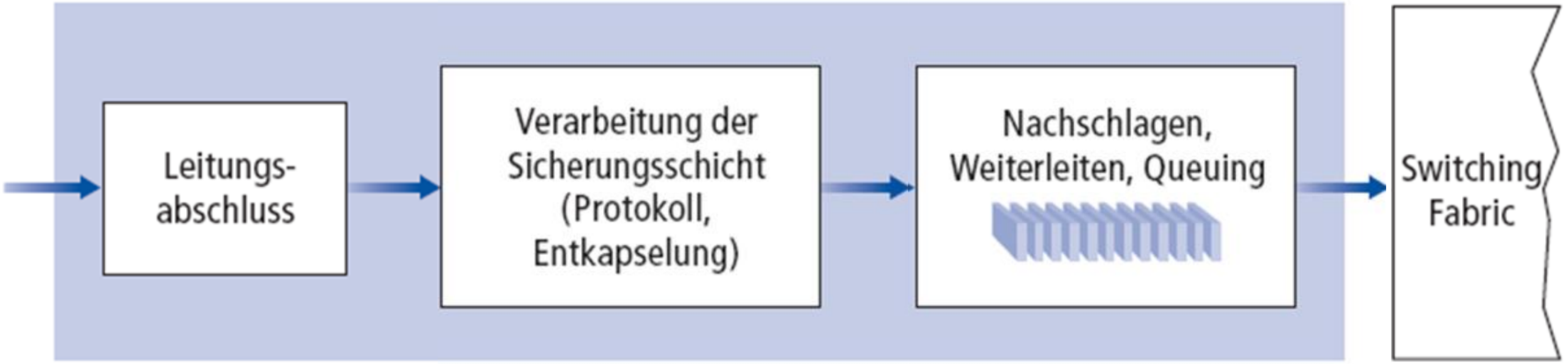


Übersicht: Routerarchitektur

Zwei wichtige Aufgaben eines Routers:

- Ausführen von Routing-Algorithmen und -Protokollen
    - RIP, OSPF, BGP
- Weiterleiten von Datagrammen von einem eingehenden zu einem ausgehenden Link

Verarbeitung im Eingangsport

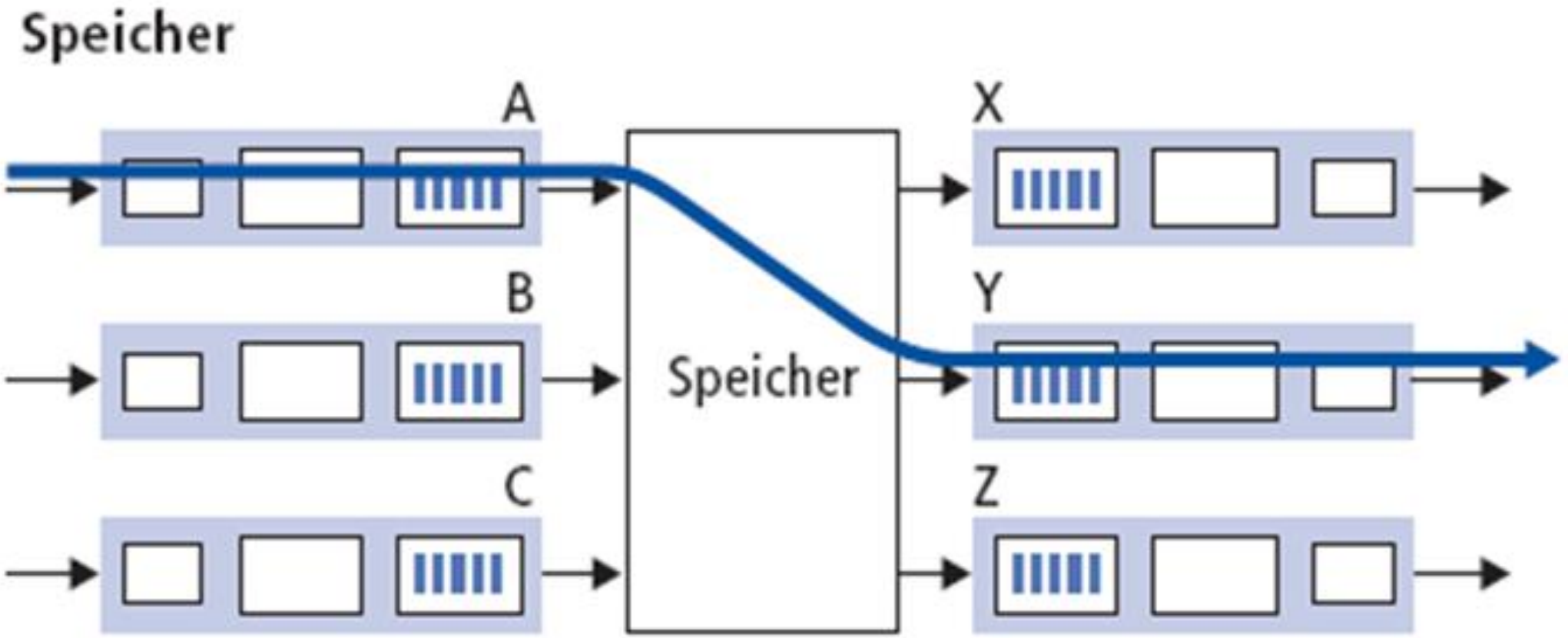


physikalische Schicht:
Bits empfangen

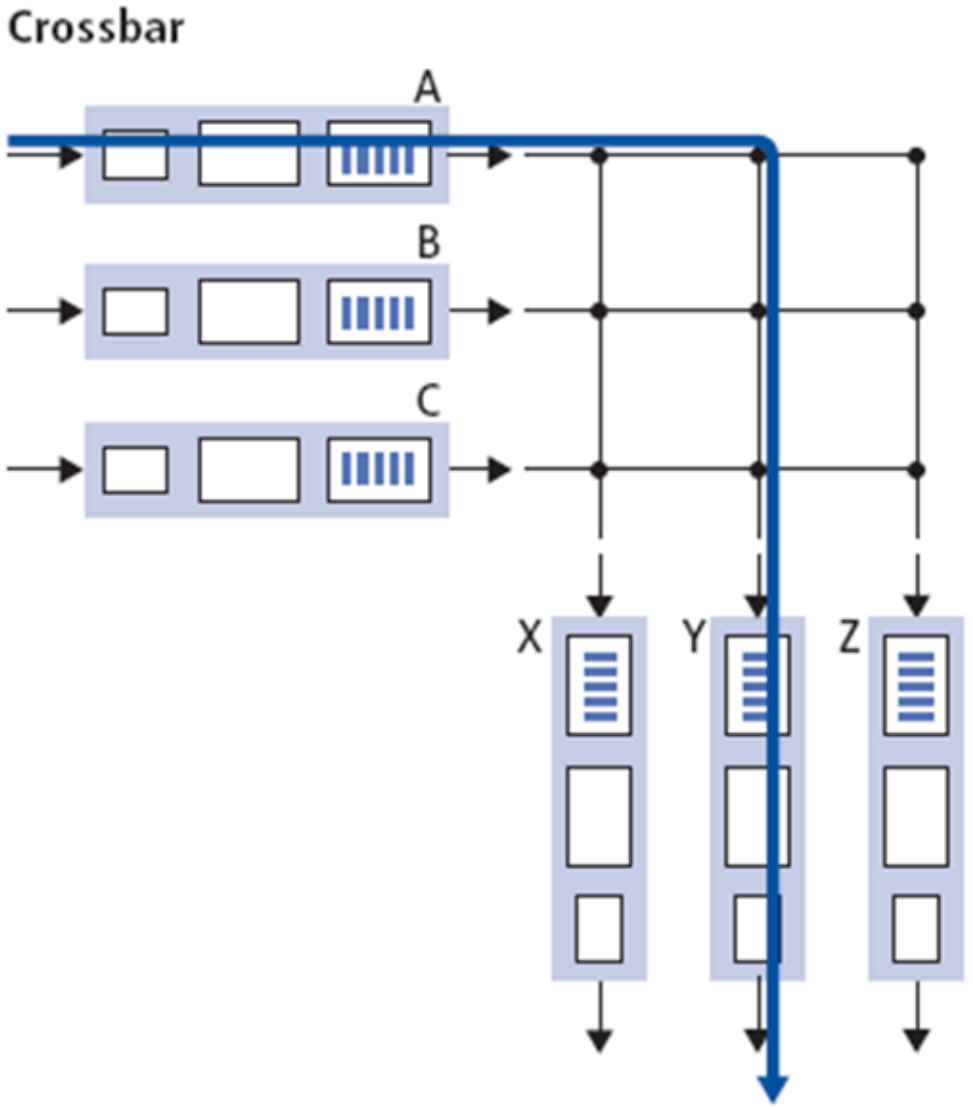
Sicherungsschicht:
z..B. Ethernet

- Suche nach einem geeigneten Ausgangsport
- Dezentral, Kopie der Routing-Tabelle (oder Teile davon) notwendig
- Ziel: Behandlung der Pakete mit „line speed“, also mit der Geschwindigkeit der Eingangsleitung des Ports
- Puffern von Paketen, wenn die Switching Fabric belegt ist

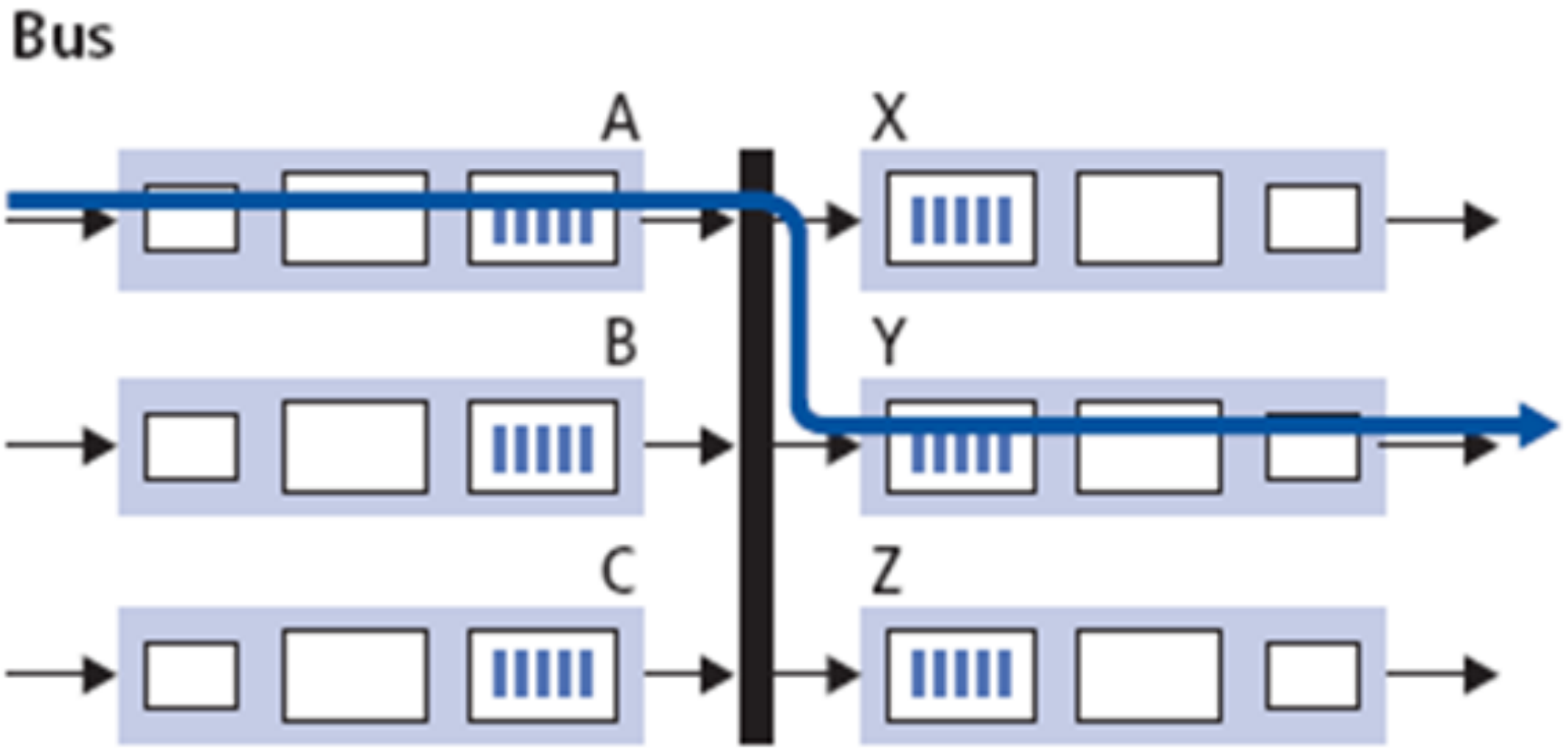
Switching Fabric



Switching Fabric



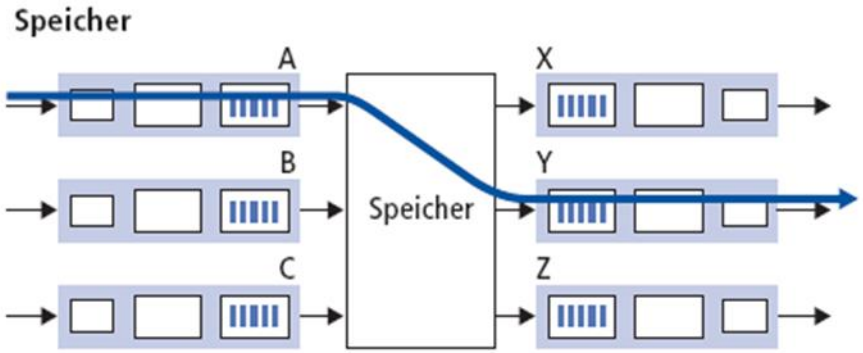
Switching Fabric



Switching über den Speicher

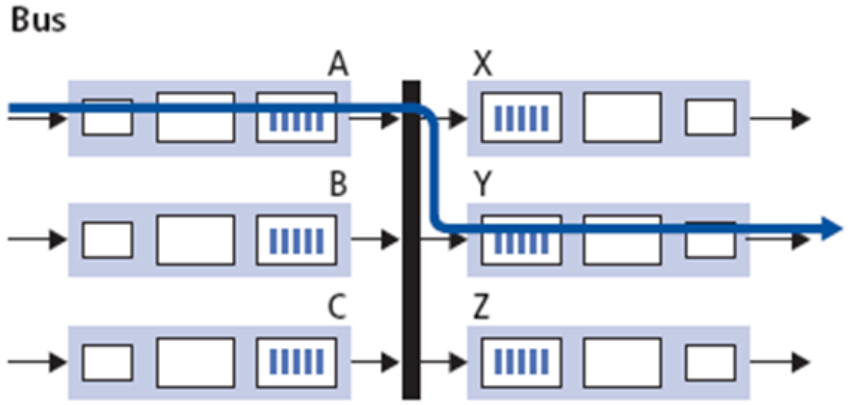
Erste Routergeneration:

- „Normale“ Rechner, Switching wird über die CPU durchgeführt
- Paket von Eingangsport in den Hauptspeicher kopieren
- Paket vom Hauptspeicher in den Ausgangsport kopieren
- Geschwindigkeit durch Speicherbus beschränkt!
    - **Zwei Speicherzugriffe: einer zum Schreiben, einer zum Lesen**



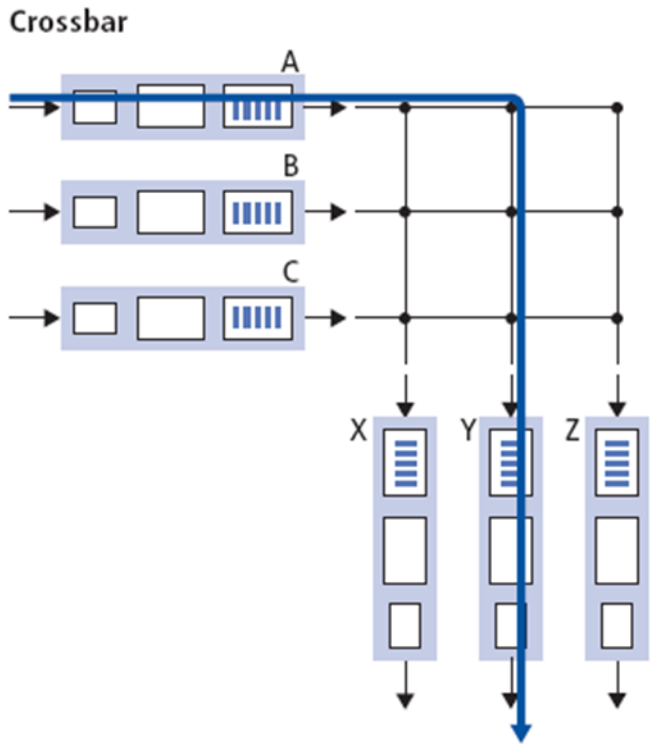
Switching über einen Bus

- Alle Ports teilen sich einen gemeinsamen Bus
- Bus Contention: Die gesamte Kommunikation erfolgt über den Bus, dieser beschränkt die Bandbreite des Routers
    - **Aber: nur eine Busoperation (nicht zwei!)**
- Beispiel: 32-Gbps-Bus, Cisco 5600, ausreichend für Zugangsrouter und Router für Firmennetze (nicht geeignet im Backbone)

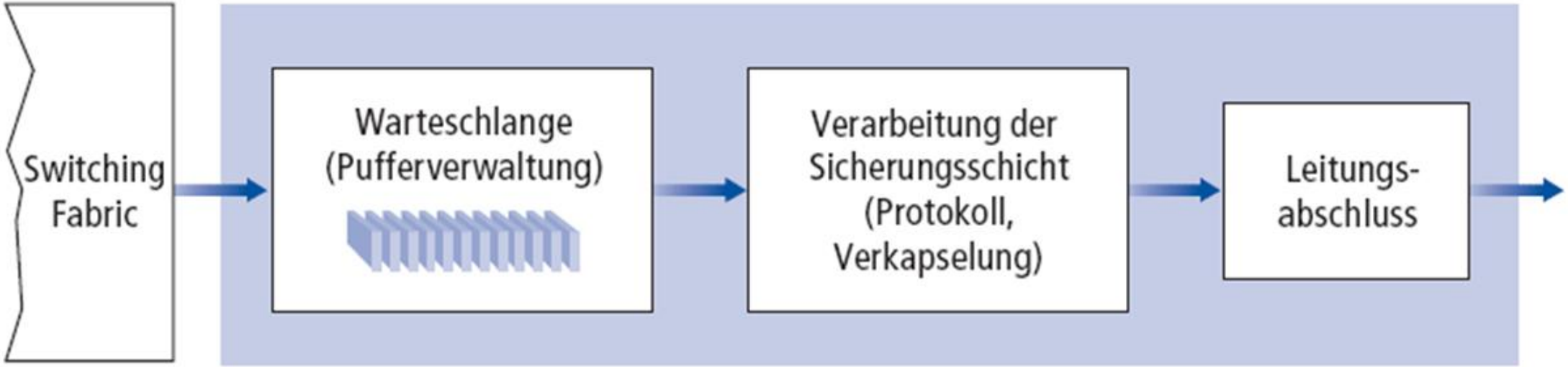


Switching über ein Spezialnetz

- Ports sind über ein Netzwerk miteinander verbunden
  - Beispielsweise alle Eingangsports über einen Crossbar mit allen Ausgangs-ports
  - Oder Banyan-Netzwerke
  - Technologie ursprünglich für das Verbinden mehrerer Prozessoren in einem Parallelrechner entwickelt
- Weitere Fortschritte: Zerlegen der Pakete in Zellen fester Größe, Zellen können dann schneller durch die Switching Fabric geleitet werden
- Beispiel: Cisco 12000, Switching von 60 Gbps durch das interne Netz



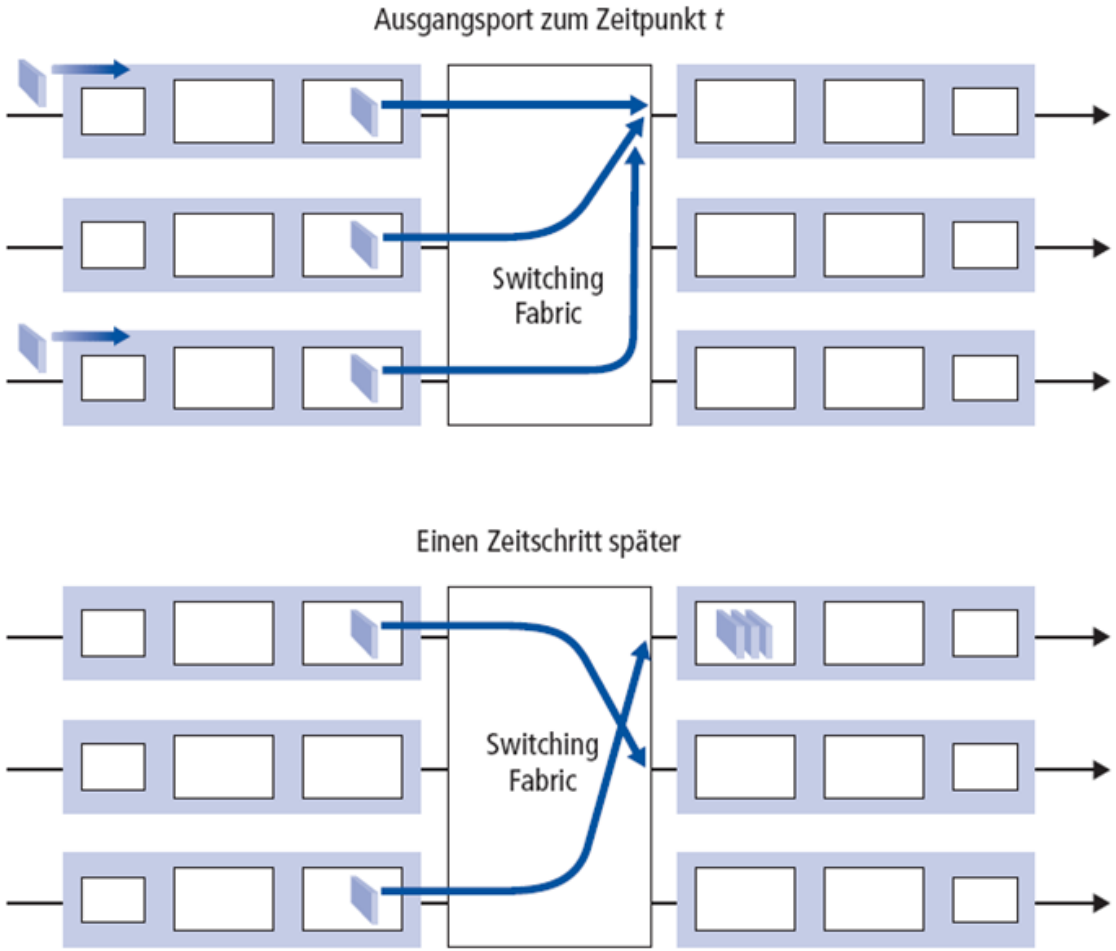
Verarbeitung im Ausgangsport



- Prinzipiell: analog zum Eingangsport!
- Einfacher, da die Entscheidung über die Weiterleitung schon getroffen ist

Puffern im Ausgangsport

- Puffern von Paketen, wenn sie schneller aus der Switching Fabric kommen, als sie auf die Leitung gelegt werden können
- Auswirkungen:
  - Gepufferte Pakete werden verzögert
  - Wenn der Puffer überläuft, müssen Pakete verworfen werden
- „Scheduling Discipline“: bestimmt die Reihenfolge, in der gepufferte Pakete auf die Leitung gelegt werden



Puffergröße

- RFC 3439 beschreibt folgende Faustregel: Die Größe des Puffers sollte der Rundlauf-zeit (RTT, z.B. 250 ms) multipliziert mit der Datenrate des Links entsprechen
    - 10 Gps Link, 250 ms RTT, ergibt 2,5 Gbit Puffer
- Neuere Empfehlungen: bei N Datenflüssen und Link-Datenrate C:

$$\frac{RTT \cdot C}{\sqrt{N}}$$

Puffern im Inputport

- Wenn die Switching Fabric ein Paket nicht direkt weiterleiten kann, muss dieses im Eingangsport gepuffert werden
- Dort kann es ein Paket blockieren, welches eigentlich bereits durch die Switching Fabric geleitet werden könnte
    - Head-of-Line (HOL) Blocking