Systeme II

8. Vorlesungswoche
16.06. – 20.06.2008

Albert-Ludwigs-Universität Freiburg
Institut für Informatik
Rechnernetze und Telematik
Christian Schindelhauer
Sommer 2008

1

Systeme II

Kapitel 5
Vermittlungsschicht

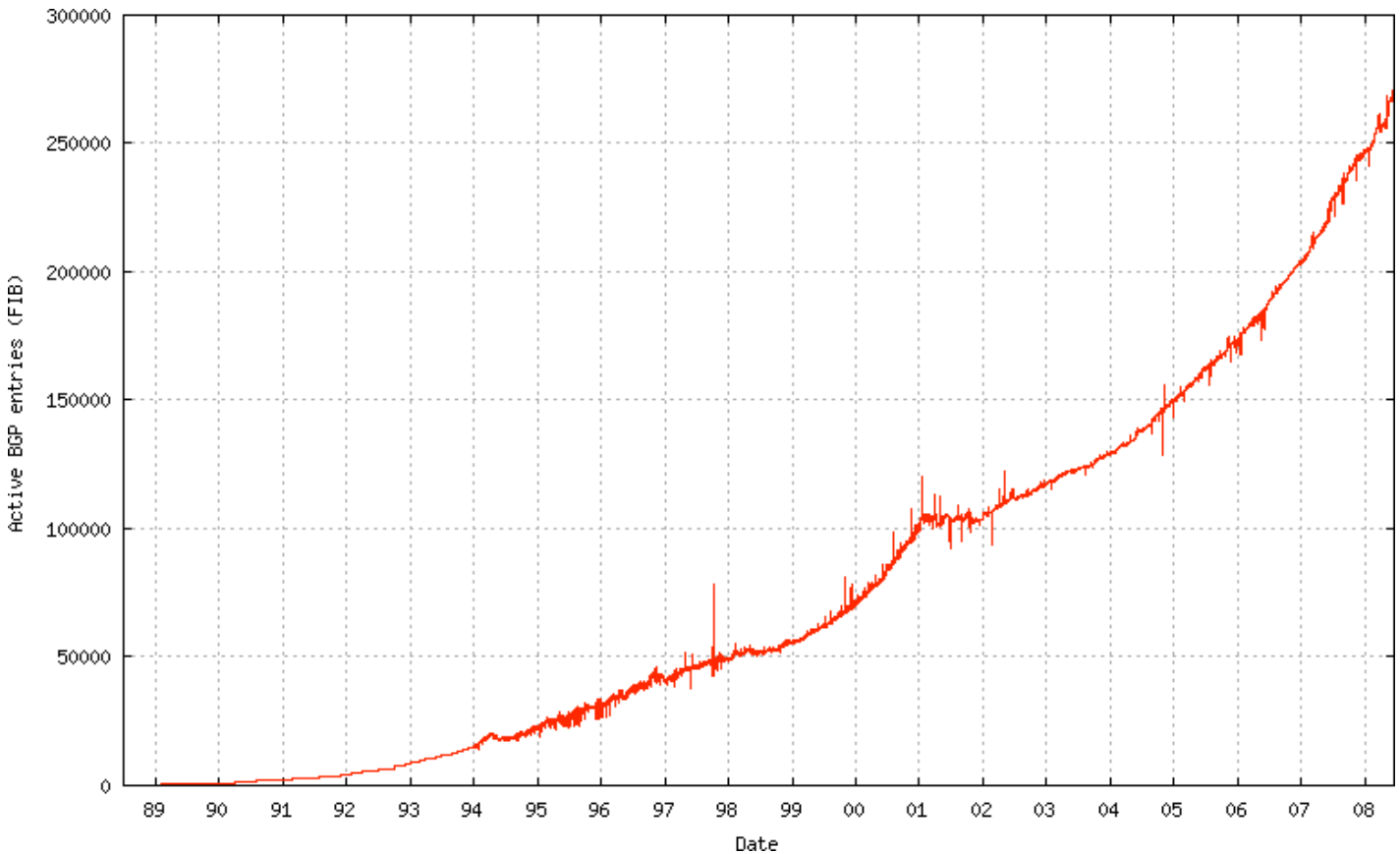
Inter-AS-Routing

- **Inter-AS-Routing ist schwierig...**
  - Organisationen können Durchleitung von Nachrichten verweigern
  - Politische Anforderungen
    - Weiterleitung durch andere Länder?
  - Routing-Metriken der verschiedenen autonomen Systeme sind oftmals unvergleichbar
    - Wegeoptimierung unmöglich!
    - Inter-AS-Routing versucht wenigstens Erreichbarkeit der Knoten zu ermöglichen
  - Größe: momentan müssen Inter-Domain-Router mehr als 250.000 Netzwerke kennen

Inter-AS: BGPv4 (Border Gateway Protocol)

- **Ist faktisch der Standard**
- **Path-Vector-Protocol**
  - ähnlich wie Distance Vector Protocol
    - es werden aber ganze Pfade zum Ziel gespeichert
  - jeder Border Gateway teilt all seinen Nachbarn (peers) den gesamten Pfad (Folge von ASen) zum Ziel mit (advertisement) (per TCP)
- **Falls Gateway X den Pfad zum Peer-Gateway W sendet**
  - dann kann W den Pfad wählen oder auch nicht
  - Optimierungskriterien:
    - Kosten, Politik, etc.
  - Falls W den Pfad von X wählt, dann publiziert er
    - Path(W,Z) = (W, Path (X,Z))
- **Anmerkung**
  - X kann den eingehenden Verkehr kontrollieren durch Senden von Advertisements
  - Sehr kompliziertes Protokoll

BGP-Routing Tabellengröße 1989-2008



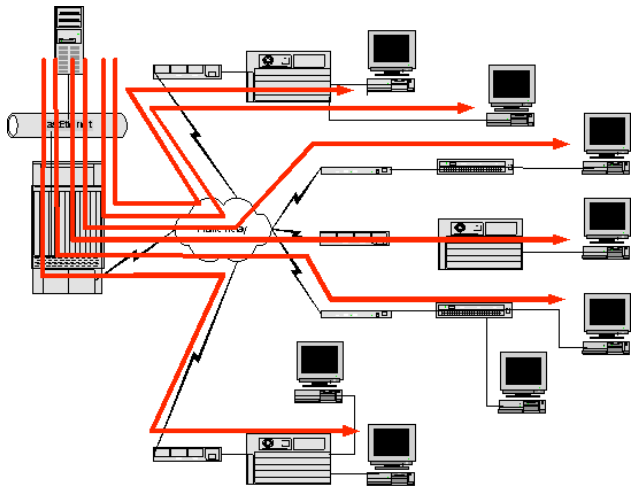
http://bgp.potaroo.net/as1221/bgp-active.html (16.06.08)

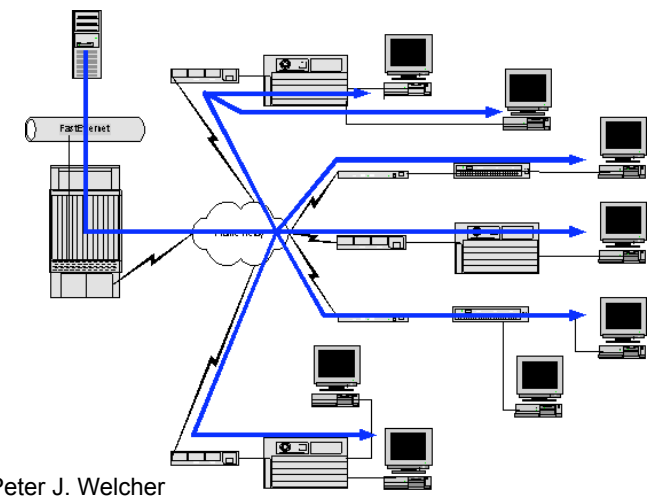
Broadcast & Multicast

- **Broadcast routing**
  - Ein Paket soll (in Kopie) an alle ausgeliefert werden
  - Lösungen:
    - Fluten des Netzwerks
    - Besser: Konstruktion eines minimalen Spannbaums
- **Multicast routing**
  - Ein Paket soll an eine gegebene Teilmenge der Knoten ausgeliefert werden (in Kopie)
  - Lösung:
    - Optimal: Steiner Baum Problem (bis heute nicht lösbar)
    - Andere (nicht-optimale) Baum-konstruktionen

IP Multicast

- **Motivation**
  - Übertragung eines Stroms an viele Empfänger
- **Unicast**
  - Strom muss mehrfach einzeln übertragen werden
  - Bottleneck am Sender
- **Multicast**
  - Strom wird über die Router vervielfältigt
  - Kein Bottleneck mehr





Bilder von Peter J. Welcher
www.netcraftsmen.net/.../ papers/multicast01.html

Funktionsprinzip

- **IPv4 Multicast-Adressen**
  - in der Klasse D (außerhalb des CIDR - Classless Interdomain Routings)
  - 224.0.0.0 - 239.255.255.255
- **Hosts melden sich per IGMP bei der Adresse an**
  - IGMP = Internet Group Management Protocol
  - Nach der Anmeldung wird der Multicast-Tree aktualisiert
- **Source sendet an die Multicast-Adresse**
  - Router duplizieren die Nachrichten an den Routern
  - und verteilen sie in die Bäume
- **Angemeldete Hosts erhalten diese Nachrichten**
  - bis zu einem Time-Out
  - oder bis sie sich abmelden
- **Achtung:**
  - Kein TCP, nur UDP
  - Viele Router lehnen die Beförderung von Multicast-Nachrichten ab
    - Lösung: Tunneln

Multicast Routing Protokolle

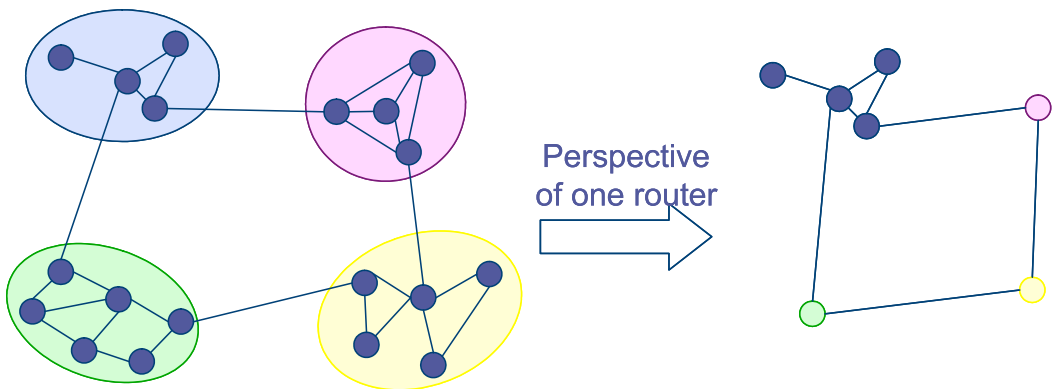
- **Distance Vector Multicast Routing Protocol (DVMRP)**
  - jahrelang eingesetzt in MBONE (insbesondere in Freiburg)
  - Eigene Routing-Tabelle für Multicast
- **Protocol Independent Multicast (PIM)**
  - im Sparse Mode (PIM-SM)
  - aktueller Standard
  - beschneidet den Multicast Baum
  - benutzt Unicast-Routing-Tabellen
  - ist damit weitestgehend protokollunabhängig
- **Voraussetzung PIM-SM:**
  - benötigt Rendevous-Point (RP) in ein-Hop-Entfernung
  - RP muss PIM-SM unterstützen
  - oder Tunneling zu einem Proxy in der Nähe eines RP

Warum so wenig IP Multicast?

- **Trotz erfolgreichem Einsatz**
  - in Video-Übertragung von IETF-Meetings
  - MBONE (Multicast Backbone)
- **gibt es wenige ISP die IP Multicast in den Routern unterstützen.**
- **Zusätzlicher Wartungsaufwand**
  - Schwierig zu konfigurieren
  - Verschiedene Protokolle
- **Gefahr von Denial-of-Service-Attacken**
  - Implikationen größer als bei Unicast
- **Transportprotokoll**
  - Nur UDP einsetzbar
  - Zuverlässige Protokolle
    - Vorwärtsfehlerkorrektur
    - Oder proprietäre Protokolle in den Routern (z.B. CISCO)
- **Marktsituation**
  - Endkunden fragen kaum Multicast nach (benutzen lieber P2P-Netzwerke)
  - Wegen einzelner Dateien und wenigen Abnehmern erscheint ein Multicast wenig erstrebenswert (Adressenknappheit!)

Adressierung und Hierarchisches Routing

- **Flache (MAC-) Adressen haben keine Strukturinformation**



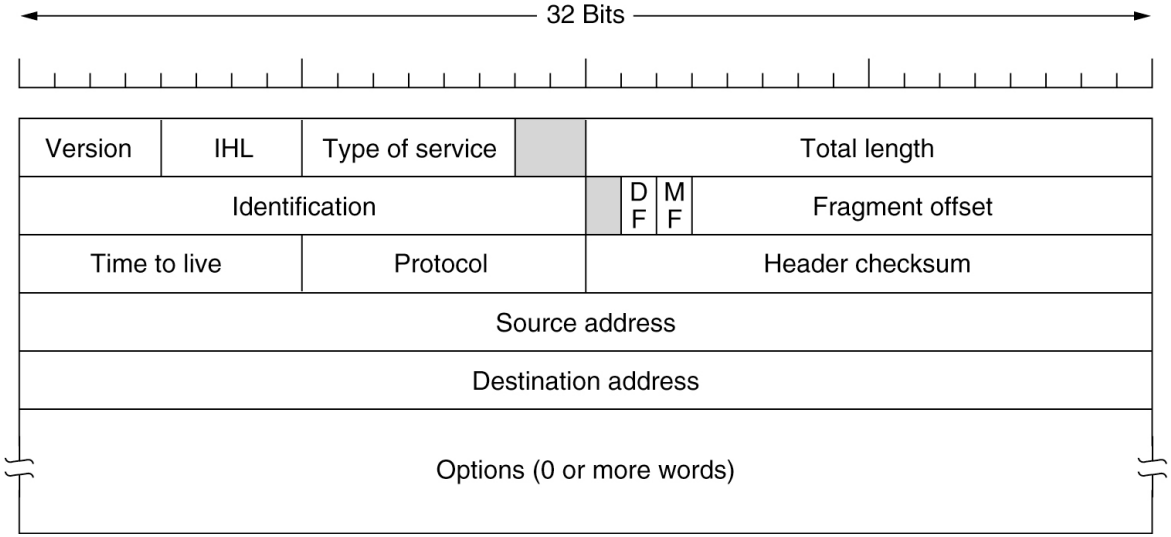
- **Hierarchische Adressen**
  - Routing wird vereinfacht wenn Adressen hierarchische Routing-Struktur abbilden
  - Group-ID$_n$:Group-ID$_{n-1}$:…:Group-ID$_1$:Device-ID

IP-Adressen und Domain Name System

- **IP-Adressen**
  - Jedes Interface in einem Netzwerk hat weltweit eindeutige IP-Adresse
  - 32 Bits unterteilt in Net-ID und Host-ID
  - Net-ID vergeben durch Internet Network Information Center
  - Host-ID durch lokale Netzwerkadministration
- **Domain Name System (DNS)**
  - Ersetzt IP-Adressen wie z.B. 132.230.167.230 durch Namen wie z.B. falcon.informatik.uni-freiburg.de und umgekehrt
  - Verteilte robuste Datenbank

IPv4-Header (RFC 791)

- **Version: 4 = IPv4**
- **IHL: IP Headerlänge**
  - in 32 Bit-Wörtern (>5)
- **Type of Service**
  - Optimiere delay, throughput, reliability, monetary cost



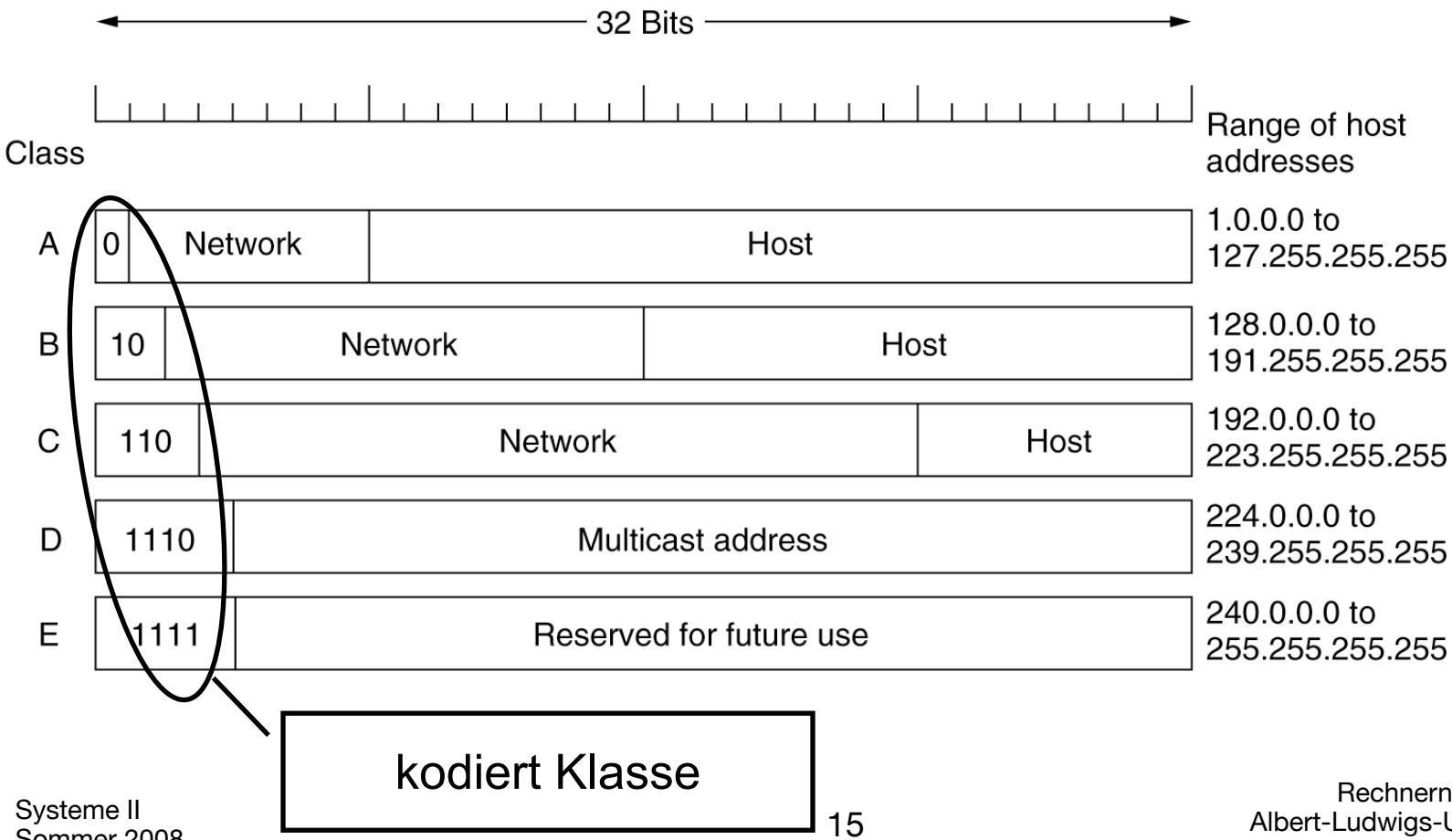
- **Checksum (nur für IP-Header)**
- **Source and destination IP-address**
- **Protocol, identifiziert passendes Protokoll**
  - Z.B. TCP, UDP, ICMP, IGMP
- **Time to Live:**
  - maximale Anzahl Hops

Internet IP Adressen bis 1993

- **IP-Adressen unterscheiden zwei Hierarchien**
  - Netzwerk-Interfaces
  - Netzwerke
    - Verschiedene Netzwerkgrößen
    - Netzwerkklassen:
      - Groß - mittel - klein
        (Klasse A, B, and C)
- **Eine IP-Adresse hat 32 Bits**
  - Erster Teil: Netzwerkadresse
  - Zweiter Teil: Interface

IP-Klassen bis 1993

- **Klassen A, B, and C**
- **D für multicast; E: “reserved”**



128 NWs; 16 M hosts

16K NWs; 64K hosts

2M NWs; 256 hosts

Rechnernetze und Telematik
Albert-Ludwigs-Universität Freiburg
Christian Schindelhauer

IPv4-Adressen

- **Bis 1993 (heutzutage veraltet)**
  - 5 Klassen gekennzeichnet durch Präfix
  - Dann Subnetzpräfix fester Länge und Host-ID (Geräteteil)
- **Seit 1993**
  - Classless Inter-Domain-Routing (CIDR)
  - Die Netzwerk-Adresse und die Host-ID (Geräteteil) werden variabel durch die Netzwerkmaske aufgeteilt.
  - Z.B.:
    - Die Netzwerkmaske 11111111.11111111.11111111. 00000000
    - Besagt, dass die IP-Adresse
      - 10000100. 11100110. 10010110. 11110011
      - Aus dem Netzwerk 10000100. 11100110. 10010110
      - den Host 11110011 bezeichnet
- **Route aggregation**
  - Die Routing-Protokolle BGP, RIP v2 und OSPF können verschiedene Netzwerke unter einer ID anbieten
    - Z.B. alle Netzwerke mit Präfix 10010101010* werden über Host X erreicht

Umwandlung in MAC-Adressen: ARP

- **Address Resolution Protocol (ARP)**
- **Umwandlung: IP-Adresse in MAC-Adresse**
  - Broadcast im LAN, um nach Rechner mit passender IP-Adresse zu fragen
  - Knoten antwortet mit MAC-Adresse
  - Router kann dann das Paket dorthin ausliefern

IPv6

Wozu IPv6:

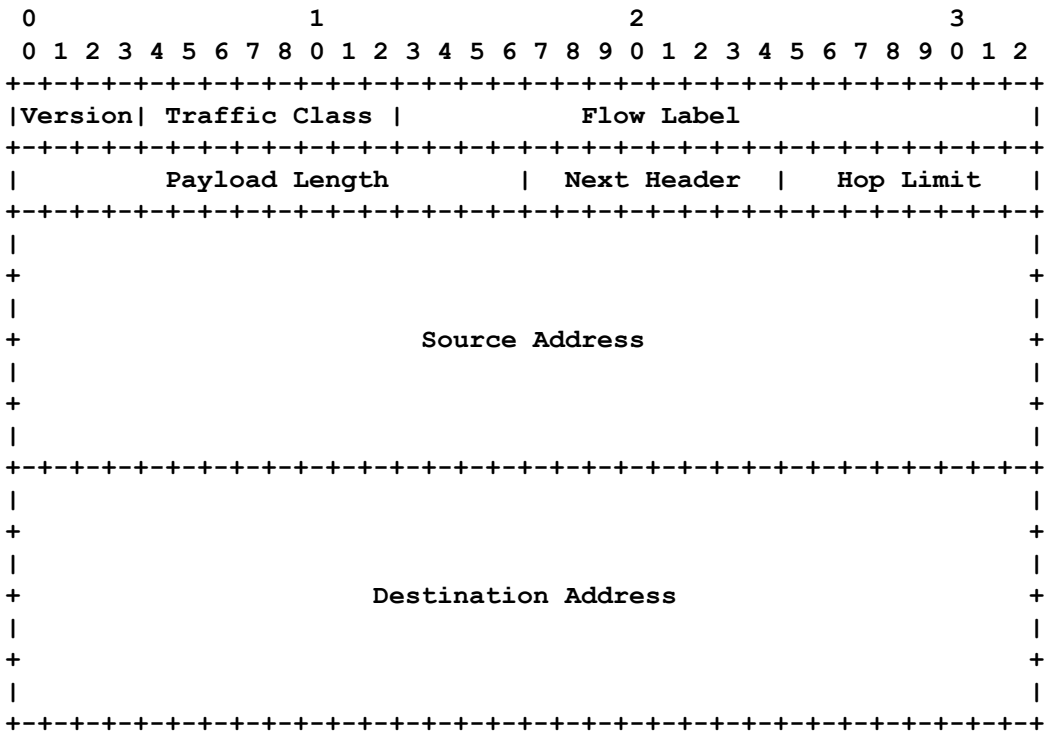
- **IP-Adressen sind knapp**
  - Zwar gibt es 4 Milliarden in IPv4 (32 Bit)
  - Diese sind aber statisch organisiert in Netzwerk- und Rechnerteil
    - Adressen für Funktelefone, Kühlschränke, Autos, Tastaturen, etc...
- **Autokonfiguration**
  - DHCP, Mobile IP, Umnummerierung
- **Neue Dienste**
  - Sicherheit (IPSec)
  - Qualitätssicherung (QoS)
  - Multicast
- **Vereinfachungen für Router**
  - keine IP-Prüfsummen
  - Keine Partitionierung von IP-Paketen

Lösung der Adressenknappheit: DHCP

- **DHCP (Dynamic Host Configuration Protocol)**
  - Manuelle Zuordnung (Bindung an die MAC-Adresse, z.B. für Server)
  - Automatische Zuordnung (feste Zuordnung, nicht voreingestellt)
  - Dynamische Zuordnung (Neuvergabe möglich)
- **Einbindung neuer Rechner ohne Konfiguration**
  - Rechner „holt“ sich die IP-Adresse von einem DHCP-Server
  - Dieser weist dem Rechner die IP-Adressen dynamisch zu
  - Nachdem der Rechner das Netzwerk verlässt, kann die IP-Adresse wieder vergeben werden
  - Bei dynamischer Zuordnung, müssen IP-Adressen auch „aufgefrischt“ werden
  - Versucht ein Rechner eine alte IP-Adresse zu verwenden,
    - die abgelaufen ist oder
    - schon neu vergeben ist
  - Dann werden entsprechende Anfragen zurückgewiesen
  - Problem: Stehlen von IP-Adressen

IPv6-Header (RFC 2460)

- **Version: 6 = IPv6**
- **Traffic Class**
  - Für QoS (Prioritätsvergabe)
- **Flow Label**
  - Für QoS oder Echtzeitanwendungen
- **Payload Length**
  - Größe des Rests des IP-Pakets (Datagramms)
- **Next Header (wie bei IPv4: protocol)**
  - Z.B. ICMP, IGMP, TCP, EGP, UDP, Multiplexing, ...
- **Hop Limit (Time to Live)**
  - maximale Anzahl Hops
- **Source Address**
- **Destination Address**
  - 128 Bit IPv6-Adresse

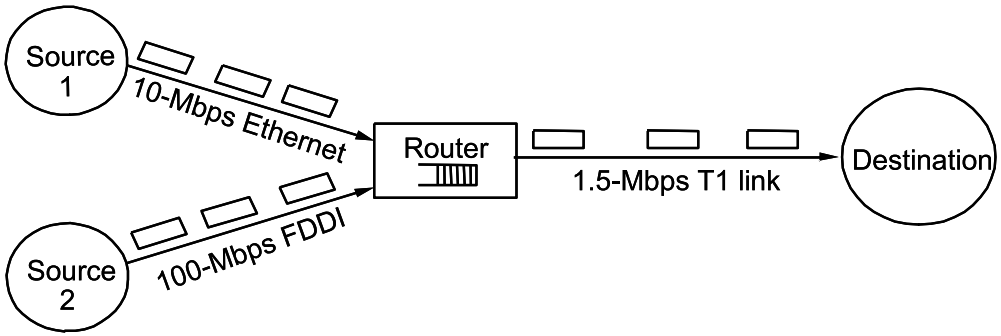


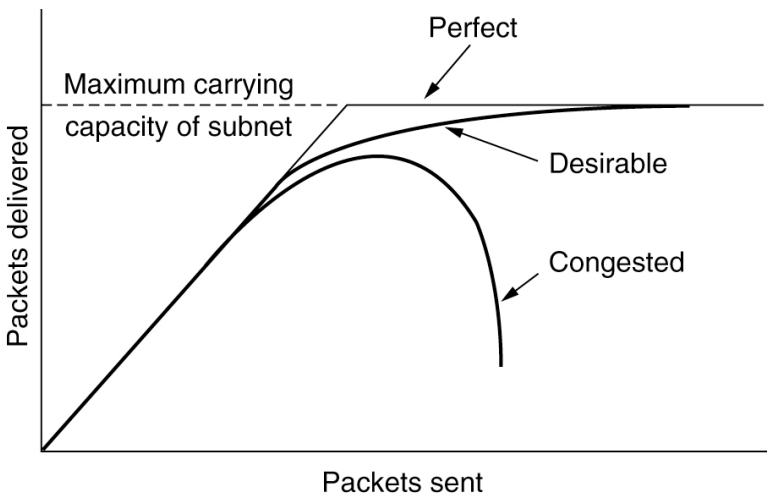
IPsec (RFC 2401)

- **Schutz vor Replay-Attacken**
- **IKE (Internet Key Exchange) Protokoll**
  - Vereinbarung einer Security Association
    - Identifikation, Festlegung von Schlüsseln, Netzwerke, Erneuerungszeiträume für Authentifizierung und IPsec Schlüssel
  - Erzeugung einer SA im Schnellmodus (nach Etablierung)
- **Encapsulating Security Payload (ESP)**
  - IP-Kopf unverschlüsselt, Nutzdaten verschlüsselt, mit Authentifizierung
- **IPsec im Transportmodus (für direkte Verbindungen)**
  - IPsec Header zwischen IP-Header und Nutzdaten
  - Überprüfung in den IP-Routern (dort muss IPsec vorhanden sein)
- **IPsec im Tunnelmodus (falls mindestens ein Router dazwischen ist)**
  - Das komplette IP-Paket wird verschlüsselt und mit dem IPsec-Header in einen neuen IP-Header verpackt
  - Nur an den Enden muss IPsec vorhanden sein.
- **IPsec ist Bestandteil von IPv6**
- **Rückportierungen nach IPv4 existieren**

Congestion Control
Stauvermeidung

- **Jedes Netzwerk hat eine eingeschränkte Übertragungs-Bandbreite**
- **Wenn mehr Daten in das Netzwerk eingeleitet werden, führt das zum**
  - Datenstau (congestion) oder gar
  - Netzwerkzusammenbruch (congestive collapse)
- **Folge: Datenpakete werden nicht ausgeliefert**





Schneeballeffekt

- **Congestion control soll Schneeballeffekte vermeiden**
  - Netzwerküberlast führt zu Paketverlust (Pufferüberlauf, ...)
  - Paketverlust führt zu Neuversand
  - Neuversand erhöht Netzwerklast
  - Höherer Paketverlust
  - Mehr neu versandte Pakete
  - ...

Anforderungen an Congestion Control

- **Effizienz**
  - Verzögerung klein
  - Durchsatz hoch

- **Fairness**
  - Jeder Fluss bekommt einen fairen Anteil
  - Priorisierung möglich
    - gemäß Anwendung
    - und Bedarf

Mittel der Stauvermeidung

- **Erhöhung der Kapazität**
  - Aktivierung weiterer Verbindungen, Router
  - Benötigt Zeit und in der Regel den Eingriff der Systemadministration
- **Reservierung und Zugangskontrolle**
  - Verhinderung neuen Verkehrs an der Kapazitätsgrenze
  - Typisch für (Virtual) Circuit Switching
- **Verringerung und Steuerung der Last**
  - (Dezentrale) Verringerung der angeforderten Last bestehender Verbindungen
  - Benötigt Feedback aus dem Netzwerk
  - Typisch für Packet Switching
    - wird in TCP verwendet

Orte und Maße

- **Router- oder Host-orientiert**
  - Messpunkt (wo wird der Stau bemerkt)
  - Steuerung (wo werden die Entscheidungen gefällt)
  - Aktion (wo werden Maßnahmen ergriffen)
- **Fenster-basiert oder Raten-basiert**
  - Rate: x Bytes pro Sekunde
  - Fenster: siehe Fenstermechanismen in der Sicherungsschicht
    - wird im Internet verwendet

Routeraktion: Paket löschen

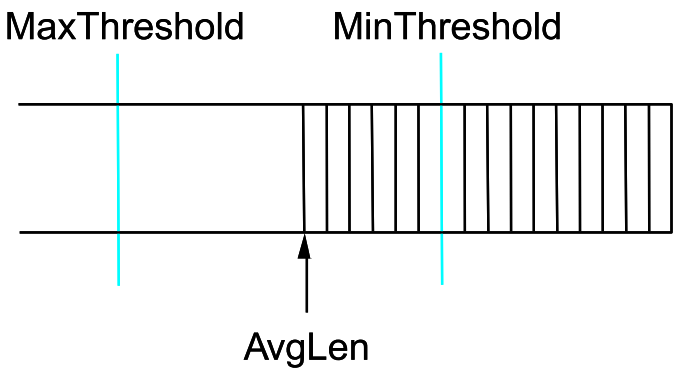
- **Bei Pufferüberlauf im Router**
  - muss (mindestens) ein Paket gelöscht werden
- **Das zuletzt angekommene Paket löschen (***drop-tail queue***)**
  - Intuition: “Alte” Pakete sind wichtiger als neue (Wein)
    - z.B. für go-back-n-Strategie
- **Ein älteres Paket im Puffer löschen**
  - Intuition: Für Multimedia-Verkehr sind neue Pakete wichtiger als alte (Milch)

Paketverlust erzeugt implizites Feedback

- **Paketverlust durch Pufferüberlauf im Router erzeugt Feedback in der Transportschicht beim Sender durch ausstehende Bestätigungen**
  - Internet
- **Annahme:**
  - Paketverlust wird hauptsächlich durch Stau ausgelöst
- **Maßnahme:**
  - Transport-Protokoll passt Senderate an die neue Situation an

Proaktive Methoden

- **Pufferüberlauf deutet auf Netzwerküberlast hin**
- **Idee: Proaktives Feedback = Stauvermeidung (Congestion avoidance)**
  - Aktion bereits bei kritischen Anzeigewerten
  - z.B. bei Überschreitung einer Puffergröße
  - z.B. wenn kontinuirlich mehr Verkehr eingeht als ausgeliefert werden kann
  - ...
  - Router ist dann in einem Warn-Zustand



Proactive Aktion:
Pakete drosseln (Choke packets)

- **Wenn der Router in dem Warnzustand ist:**
  - Sendet er Choke-Pakete (Drossel-Pakete) zum Sender
- **Choke-Pakete fordern den Sender auf die Sende-Rate zu verringern**

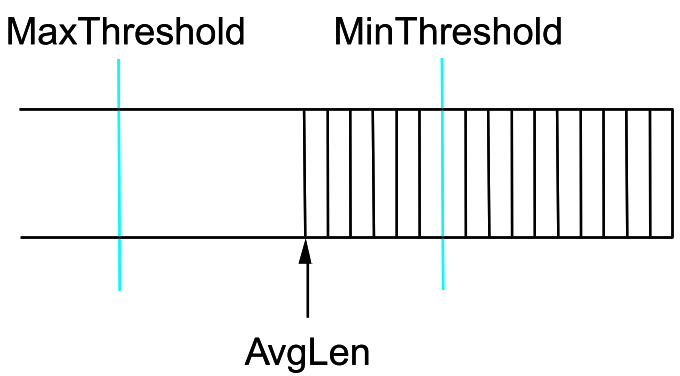
- **Problem:**
  - Im kritischen Zustand werden noch mehr Pakete erzeugt
  - Bis zur Reaktion beim Sender vergrößert sich das Problem

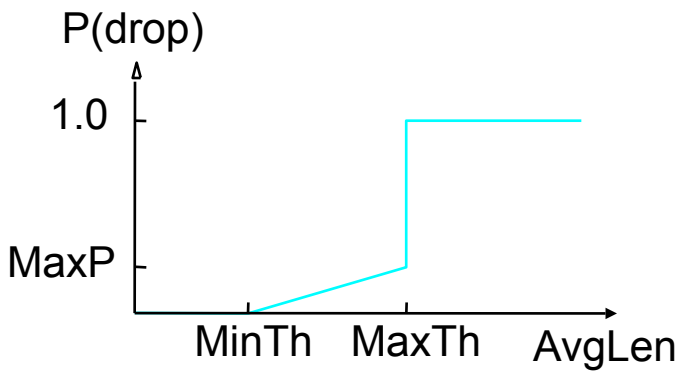
Proaktive Aktion: Warnbits

- **Wenn der Router in dem Warnzustand ist:**
  - Sendet er Warn-Bits in allen Paketen zum Ziel-Host
- **Ziel-Host sendet diese Warn-Bits in den Bestätigungs-Bits zurück zum Sender**
  - Quelle erhält Warnung und reduziert Sende-Rate

Proaktive Aktion: Random early detection (RED)

- **Verlorene Pakete werden als Indiz aufgefasst**
- **Router löschen Pakete willkürlich im Warnzustand**
- **Löschrate kann mit der Puffergröße steigen**





Reaktion des Senders

- **Raten-basierte Protokolle**
  - Reduzierung der Sende-Rate
  - Problem: Um wieviel?

- **Fenster-basierte Protokolle:**
  - Verringerung des Congestion-Fensters
  - z.B. mit AIMD (additive increase, multiplicative decrease)

Systeme II

Ende der 8. Vorlesungswoche

Albert-Ludwigs-Universität Freiburg
Institut für Informatik
Rechnernetze und Telematik
Christian Schindelhauer
Sommer 2008