

# **Internetprogrammierung Spezialvorlesung, SS2006**

**Peter Thiemann**

# Übersicht

- Einführung
- Grundlagen der Netzwerkprogrammierung in Java
- Netzwerkprotokolle der Anwendungsebene
  - DNS, HTTP, SMTP, FTP, RPC, RMI
- XML Grundlagen
  - Umfeld, Definition, DTDs, XHTML, CSS, Namespaces, XPath
- Programmierung interaktiver Webanwendungen
  - CGI, Servlets, JSP
- XML Programmierung
  - XML Parser, DOM, SAX, STX, XDuce, Xact, Xtatic

# Organisatorisches

- Übungsbetreuung durch Dipl.-Inform. Stefan Wehr
- Aufgaben in Teams zu bearbeiten
  - Programmieraufgaben
  - Experimente
  - Miniprojekt (letzte 2-3 Wochen)
- Abschlussprüfung mündlich 28./29.09. nach Terminabsprache
- Aufzeichnung vom SS2005 liegt vor
- Webseite der Vorlesung

<http://proglang.informatik.uni-freiburg.de/teaching/inetprog/2006/>

# 1 Das Internet

- globales Kommunikationssystem
- Verbindungen zwischen angeschlossenen Endgeräten
  - unicast** Rechner — Rechner
  - oder
  - multicast** ein Rechner — viele Rechner
- einheitlicher Adressraum  
(Internet-Adressen, Domainnamen)

# 1.1 Geschichte

**1969:** ARPANET (4 Hosts, 50kb/s)

**1973:** Netzwerkprotokoll TCP/IP (Vinton Cerf, Bob Kahn)  
Kommunikation zwischen Netzwerken

**1974:** Name "Internet"

**1976:** Ethernet (Bob Metcalfe)

**1979:** USENET/News

**1981:** ARPANET, CSNET (213 Hosts)

**1983:** DNS (562 Hosts)

**1988:** (56000 Hosts)

**1990:** Hypertext-System von Tim Berners-Lee (CERN)

**1992:** Gründung der Internet Society

World-Wide-Web Artikel veröffentlicht

“Surfing the Internet” (Jean Armour Polly)

(1.136.000 Hosts)

**1993:** Mosaic Browser

**1994:** Netscape Browser

W3C am MIT gegründet <http://www.w3c.org/>

erste kommerzielle Anwendungen:

Pizzabestellung, First Virtual, Shopping Malls

(3.864.000 Hosts)

**1995:** 6.642.000 Hosts

**1998:** > 30.000.000 Hosts; 6.500.000 Pings

**2000:** 72,398,092 DNS Einträge,

**2003:** 171,638,297 DNS Einträge

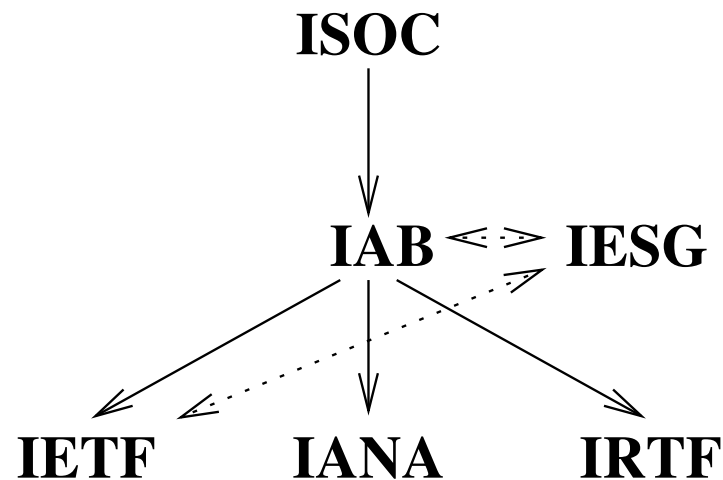
**Jan 2005:** 317,646,084 DNS Einträge aber dynamische  
(Mehrfach-) Nutzung!

**Jan 2006:** 394.991.609 IP-Adressen mit DNS Eintrag

**Quelle:** <http://www.isc.org/ds/>

## 1.2 Organisation

- dezentral und offen
- Festlegung von Protokollen und Standards durch Vorschlag, Implementierung und Bekanntmachung



**IESG** Internet Engineering Steering Group

**IANA** Internet Assigned Numbers Authority

**IRTF** Internet Research Task Force



## 1.2.1 The Internet Society

<http://www.isoc.org/>

The Internet Society is a non-profit, non-governmental, international, professional membership organization. It focuses on: standards, education, and policy issues.

## 1.2.2 Internet Architecture Board

<http://www.iab.org/iab/>

The Internet Architecture Board (IAB) is a technical advisory group of the Internet Society. Its responsibilities include:

1. **IESG Selection:** The IAB appoints a new IETF chair and all other IESG candidates, from a list provided by the IETF nominating committee.
2. **Architectural Oversight:** The IAB provides oversight of the architecture for the protocols and procedures used by the Internet.
3. **Standards Process Oversight and Appeal:** The IAB provides oversight of the process used to create Internet Standards. The IAB serves as an appeal board for complaints of improper execution of the standards process.

4. **RFC Series and IANA:** The IAB is responsible for editorial management and publication of the Request for Comments (RFC) document series, and for administration of the various Internet assigned numbers.
5. **External Liaison:** The IAB acts as representative of the interests of the Internet Society in liaison relationships with other organizations concerned with standards and other technical and organizational issues relevant to the world-wide Internet.
6. **Advice to ISOC:** The IAB acts as a source of advice and guidance to the Board of Trustees and Officers of the Internet Society concerning technical, architectural, procedural, and (where appropriate) policy matters pertaining to the Internet and its enabling technologies.

## 1.2.3 The Internet Engineering Task Force

<http://www.ietf.org/>

The Internet Engineering Task Force is a loosely self-organized group of people who make technical and other contributions to the engineering and evolution of the Internet and its technologies. **It is the principal body engaged in the development of new Internet standard specifications.**

The IETF meeting is not a conference, although there are technical presentations. The IETF is not a traditional standards organization, although many specifications are produced that become standards. The IETF is made up of volunteers who meet three times a year to fulfill the IETF mission.

IETF's mission includes:

- Identifying, and proposing solutions to, pressing operational and technical problems in the Internet;
- Specifying the development or usage of protocols and the near-term architecture to solve such technical problems for the Internet;
- Making recommendations to the Internet Engineering Steering Group (IESG) regarding the standardization of protocols and protocol usage in the Internet;
- Facilitating technology transfer from the Internet Research Task Force (IRTF) to the wider Internet community; and
- Providing a forum for the exchange of information within the Internet community between vendors, users, researchers, agency contractors and network managers.

# RFC

<http://www.ietf.org/rfc.html>

IETF gibt RFCs (Request for Comments) heraus. Diese beschreiben (meist) Protokolle (nicht Datenformate wie z.B. das W3C). Sie durchlaufen einen Standardisierungsprozess. Manche werden zu Standards.

## Anforderungen an ein RFC

- Format
- Reviewprozess: Internet-Draft  $\Rightarrow$  RFC  $\Rightarrow$  Standard (RFC Editor)
- RFCs werden nie revidiert, nur überschrieben.
- De-facto Dokumentation des Internets

## 1.3 Protokolle

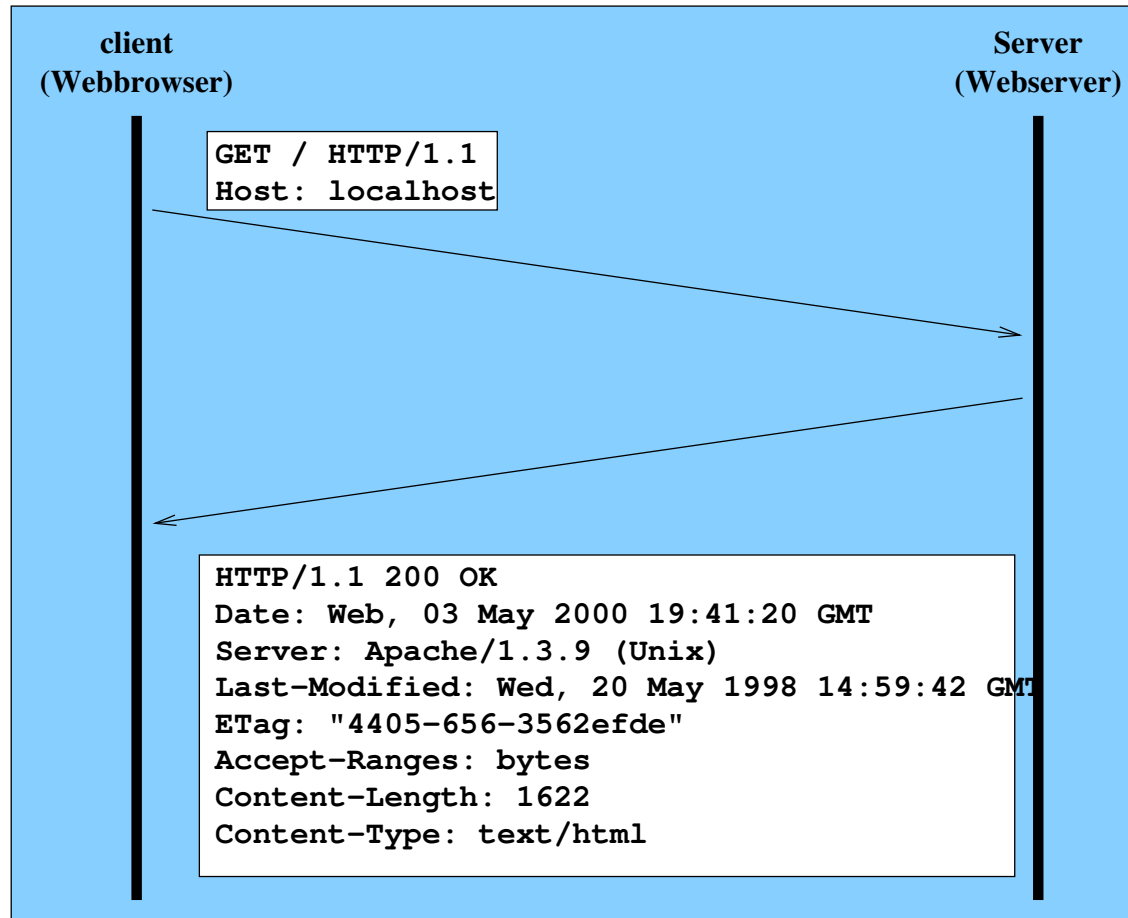
**Protokoll** Spezifikation der Struktur einer Kommunikation

**OSI-ISO Standard** definiert 7 Protokollschichten mit festgelegten Aufgaben:

1. Bitübertragung
2. Sicherung
3. Vermittlung
4. Transport
5. Kommunikation
6. Darstellung
7. Anwendung

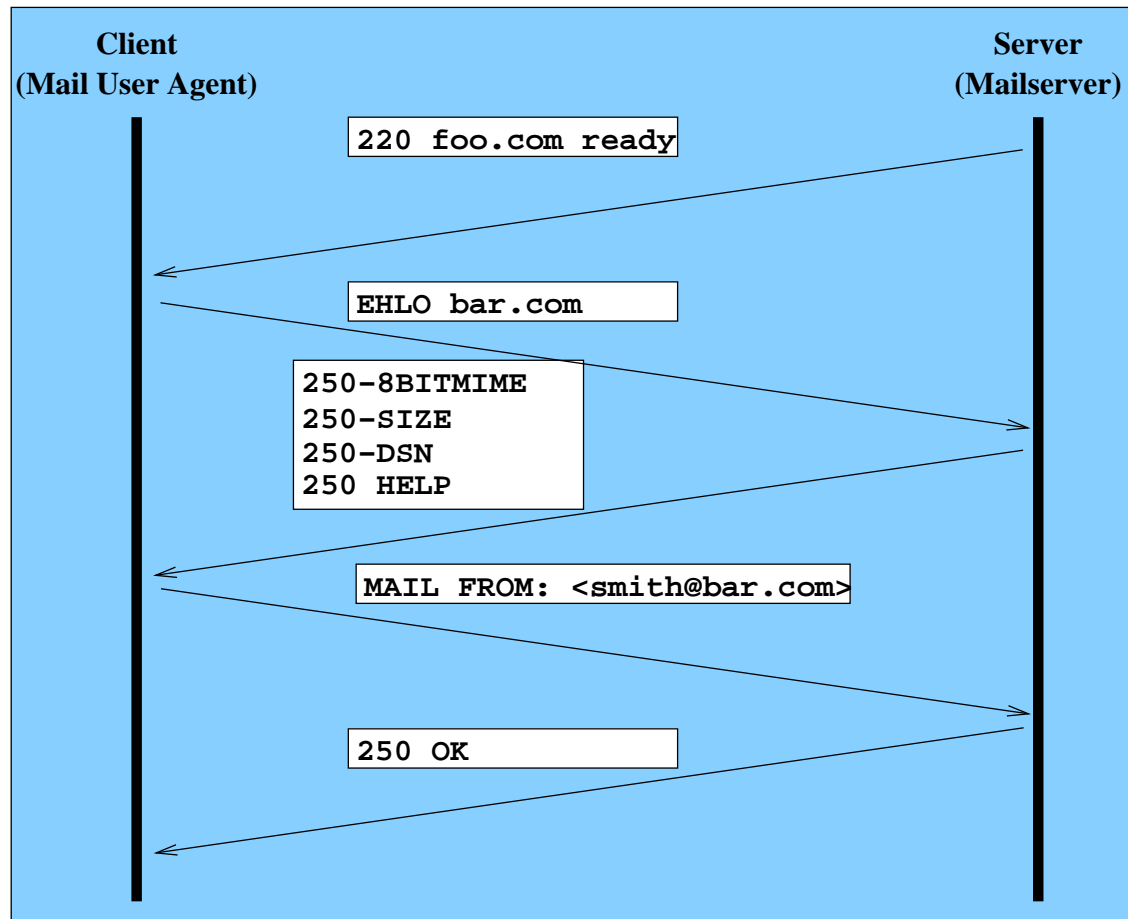
**Praktisch verwendet:** Internet Protokoll RFC 791

## 1.3.1 Beispiel: HTTP (Hypertext Transfer Protocol)





## 1.3.2 Beispiel: SMTP (Simple Mail Transfer Protocol)



## 1.3.3 IP Adresse

- Logische Rechneradresse
- Physikalische Rechneradresse wird durchs Netzwerk ermittelt (routing)
- Klassische IP Adresse (IPV4):
  - 32 Bit geschrieben in vier Dezimalzahlen  $x_1.x_2.x_3.x_4$  (je 0-255)
  - Bsp: 132.230.168.1
  - Einteilung in Klassen A, B, C, D heute irrelevant
- IPV6 Adresse:
  - 128 Bit geschrieben in acht Hexzahlen  $y_1:y_2:y_3:y_4:y_5:y_6:y_7:y_8$  mit max vier Stellen
  - Bsp: fe80::250:4ff:fe09:7028
  - Mehr Struktur, Routinginformation, etc eingebaut

## 1.3.4 Low-level Internet Protokolle

- Datenpaket = Folge von Oktetten
- Adressierung mit Hilfe von IP

### UDP – User Datagram Protocol

- Versenden eines Datenpakets (unidirektional)
- Keinerlei Garantien!

# TCP/IP – transmission control protocol / internet protocol

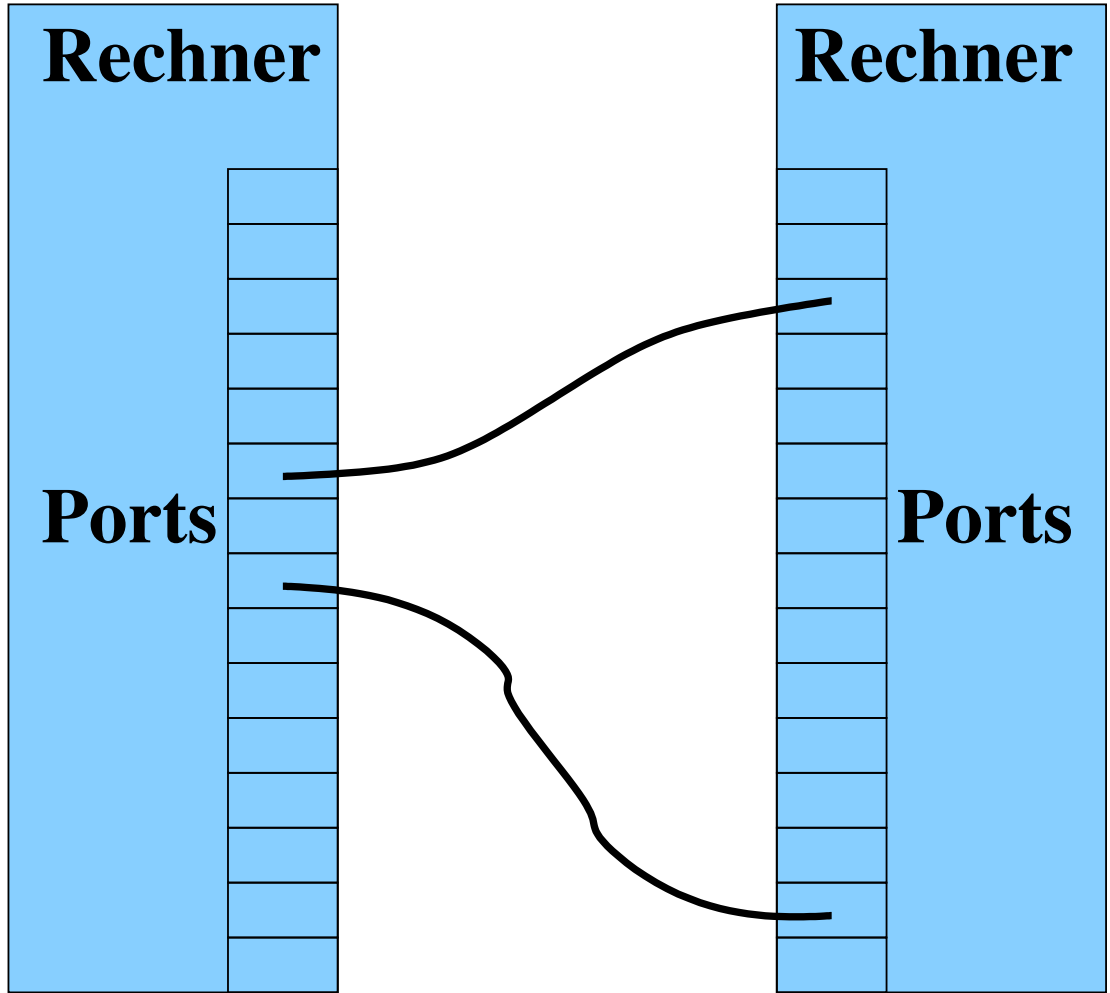
Virtuelle, strombasierte Verbindung (*stream connection*), d.h. keine feste Länge für Datenpakete vorgegeben.

- Verbindungsaufbau
- **verlässliche** bidirektionale Verbindung
  - Bündelung zu Paketen
  - Paketbestätigung
  - Paketverlust
  - Reihenfolge
  - Lastregelung
- Verbindungsabbau

## 1.3.5 Dienste (Services)

- Protokolle auf höherer Abstraktionsebene
- meist aufbauend auf TCP/IP
- meist Client-Server Struktur  
aktuell interessant: Peer-to-peer Protokolle z.B. für  
Filesharing (edonkey, gnutella, freenet, etc)

# Ein Rechner – mehrere Dienste



- Auswahl der Dienste durch *Portnummern* (16 Bit) oder Servicenamen (vgl. `/etc/services`; RFC 1700)
- Well Known Ports: 0 – 1023  
Server muss vom Administrator gestartet werden
- Registered Ports: 1024 – 49151  
Keine Restriktion bzgl. Server
- Dynamic/Private Ports: 49152 – 65535

## 1.3.6 Beispiele für Dienste

**Telnet** RFC 0854

Virtuelle Terminals für Remote Login (viele Erweiterungen, seit 1972)

**Server:** telnetd

**Client:** telnet; telnet *<Host>* *<Portnummer>*

mit Verschlüsselung: sshd bzw. ssh (mit anderem Protokoll)



**E-mail** Versendung von E-mail: SMTP (Simple Mail Transfer Protocol) RFC 2821

Format einer E-mail: RFC 2822

Erweiterte Mail Inhalte: MIME (Multipurpose Internet Mail Extensions) RFC 2045 – RFC 2049 uva

**Server:** sendmail oder ein anderer MTA (Mail Transfer Agent)

**Client:** mail oder ein anderer MUA (Mail User Agent),  
z.B. emacs, pine, netscape, mh, ...

# File Transfer Protocol (FTP)

- Transport von Dateien
- RFC 0959 und viele Erweiterungen
- seit 1971 RFC 0114
- Persönlicher Modus: Account (Name/Passwort) auf Server erforderlich
- Anonymer Modus (anonymous FTP): jeder kann zugreifen

**Server:** ftpd

**Client:** ftp, ...

# **WWW** Hypertext Transfer Protocol (HTTP 1.1)

RFC 2616

- Erweiterbares Protokoll zur Übertragung und Manipulation von getypten Dokumenten
- Adressierung der Dokumente durch URIs (Uniform Resource Identifiers)

**Server:** AOLserver, Apache, MS IIS, Jigsaw, ...

**Clients:** mozilla, opera, msie, amaya, wget, ...

## URI (Uniform Resource Identifier) RFC 1630, RFC 3986:

*This document defines the syntax used by the World-Wide Web initiative to **encode the names and addresses of objects on the Internet**. The web is considered to include objects **accessed using an extendable number of protocols**, existing, invented for the web itself, or to be invented in the future. **Access instructions** for an individual object under a given protocol are **encoded into forms of address string**. Other protocols allow the use of object names of various forms. In order to abstract the idea of a generic object, the web needs the concepts of the universal set of objects, and of the universal set of names or addresses of objects.*

## Spezielle URIs

- **Uniform Resource Locator (URL)** RFC 1738; RFC 1808, RFC 2368, RFC 3986
  - Symbolische Adresse für ein Dokument (Objekt)
  - Enthält spezifische **Zugriffsinformation**, d.h. Rechnernamen, Passwörter, etc
  - Format:  
*⟨Schema⟩:⟨schemaspezifische Information⟩*
  - Beispiele:  
`http://www.informatik.uni-freiburg.de/proglang`  
`ftp://ftp.informatik.uni-freiburg.de/iif`  
`mailto:president@whitehouse.gov`

- **Uniform Resource Name (URN) RFC 2141**

- **Eindeutiger Name** für ein Dokument (Objekt)
- Impliziert globalen Namensraum und Persistenz
- Namensraum verwaltet durch IANA  
<http://www.iana.org/assignments/urn-namespaces>

- Format:

`urn:⟨NI⟩:⟨NSS⟩`

*NI* — Namespace Identifier (assigned by IANA)

*NSS* — Namespace Specific String

- Beispiel:

`urn:ietf:rfc:2141`

`urn:ietf:std:50`

`urn:ISSN:1560-1560`

`urn:newsml:iptc.org:20001006:NewsMLv1.0:1`

`urn:newsml:reuters.com:20000206:`

`IIMFFH05643_2000-02-06_17-54-01_L06156584:1U`