max planck institut
informatik

Das Internet

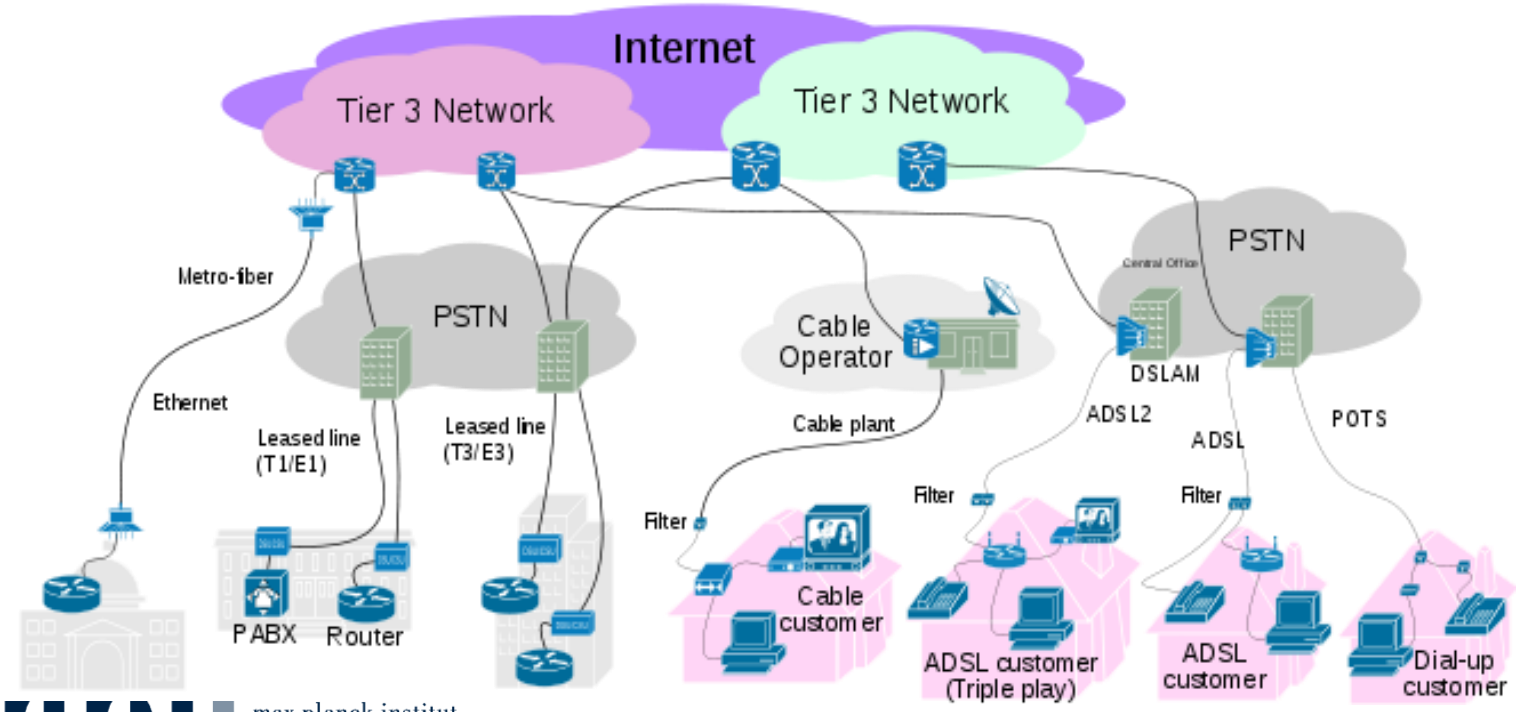
Kurt Mehlhorn
Max-Planck-Institut für Informatik

Ideen der Informatik

Viele Folien von Kosta Panagiotou

Was passiert alles,

- wenn ich eine Webseite aufrufe?
- wenn ich eine email abschicke?



Überblick

- Datenübertragung
  - zwischen zwei Rechnern
  - zwischen Rechnern in einem Netzwerk
  - zwischen Netzen im Internet
- Aufbau von Webseiten
- Darstellung im Webbrowser
- Email

Datenübertragung

- Bits werden als Spannung am Kabel übertragen, z.B. +5V = 1, -5V = 0
- …Oder per WLAN
- …Oder per Satellit
- …Oder per Brieftaube
- Unterschiede müssen für den Benutzer unsichtbar sein!

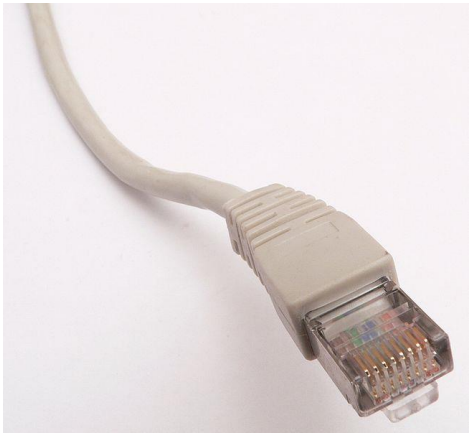
Konstruieren in Schichten

- Eine Schicht (Layer) bietet Dienste an höhere Schichten an und nutzt die Dienste der darunterliegenden Schicht zur Realisierung. Realisierung ist nach oben hin verborgen.
- Unterste Schicht setzt auf der physikalischen Realität auf.
- Klempner nutzt Rohre, Zangen, Bohrmaschine und bietet Wasserleitungen eines Hauses

Schichten

- Link Layer
  - Abstrahiert von der Technik im lokalen Netz, von der Physik zum Bit
- Internet Layer
  - Verbindet das lokale Netz mit dem Netzanbieter, Transport ohne Garantien, vom Bit zu Paketzustellung.
- Transport Layer
  - Fehlertolerante Datenübertragung.
- Data Layer
  - Kommunikationsprotokoll zwischen Browser und Server, Dienste für den Endnutzer.

Ethernet, eine populäres Netzwerk



- Kabelgebunden
- +5V = 1, -5V = 0
- 100-1000 Millionen Bits pro Sekunde

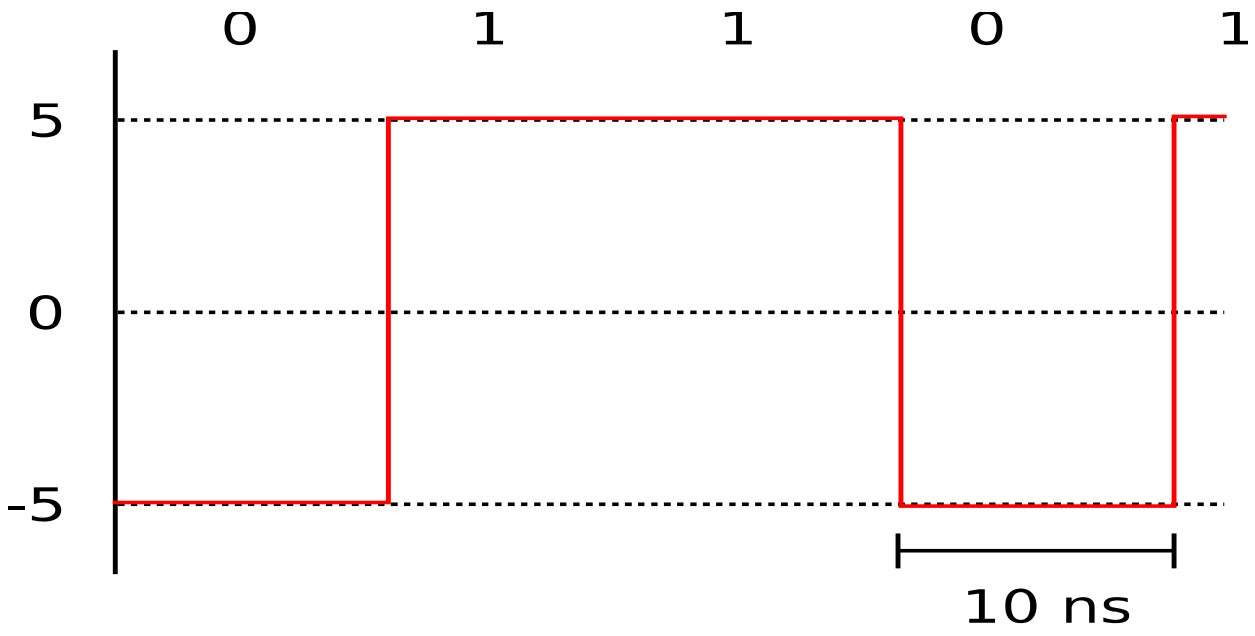


Abbildung ist stark idealisiert

Probleme

- Uhren:
  - Wann messe ich die Spannung?
  - Welche Uhrenqualität braucht man?
  - 1000000 Einsen = 10^{-2} Sekunden 5V, nicht 10^{-2} Sekunden + 10 ns
- Störungen
  - Sollte das eine 1 sein, oder hat jemand den Fön angemacht?

Selbstsynchronisierung
billige Uhren tuns auch

- Uhren mit Nanosekundenpräzision sind teuer;
- Lösung: Nie zu lange 1 oder 0 senden, z.B

Manchester-Kodierung:

  - Kodiere 0 als 01 und 1 als 10
  - Also 0001101 als 01010110100110
  - In der kodierten Folge nie mehr als 2 gleiche Symbole hintereinander; Unterscheidung von 1 und 2 Takten reicht; selbstsynchronisierend

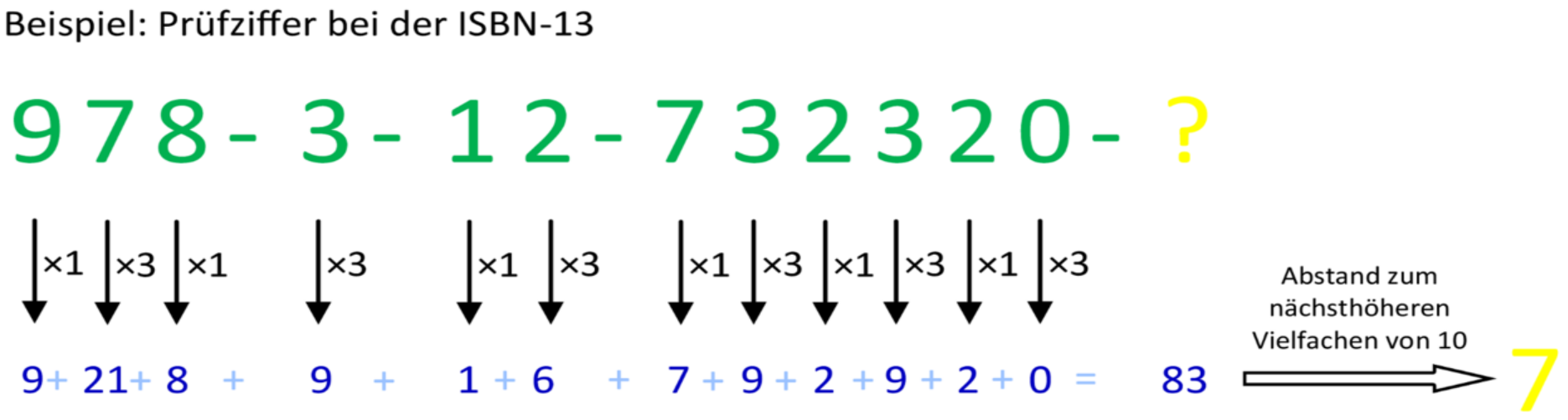
Störungen

- Übertragungsfehler passieren ständig
  - 1 Fehler pro 10 Millionen Bits = 10 Fehler/s
- Meistens: Viele Bits hintereinander falsch
- Bits werden in Pakete zusammengefasst
- Jedes Paket bekommt eine Prüfsumme; siehe nächste Folie
- Bei Fehlern im Packet: Neuübertragung

Prüfsummen

- Einfachste Prüfsumme = Quersumme
- besser (Zahlendreher): gewichtet QS

Beispiel: Prüfziffer bei der ISBN-13



Fehlerkorrigierende Kodes (Reed-Solomon)

- Ich will k ganze Zahlen senden (bei Zahlen ist die Mathematik einfacher als bei Bits), z.B. k = 128.
- Ich sende k + 2d Zahlen.
- Bis zu d Zahlen dürfen bei der Übertragung korrumpiert werden. Trotzdem kann der Empfänger die k Zahlen rekonstruieren.
- Ich zeige das Prinzip für k = 3 und d = 2.

Mathem. Hintergrund

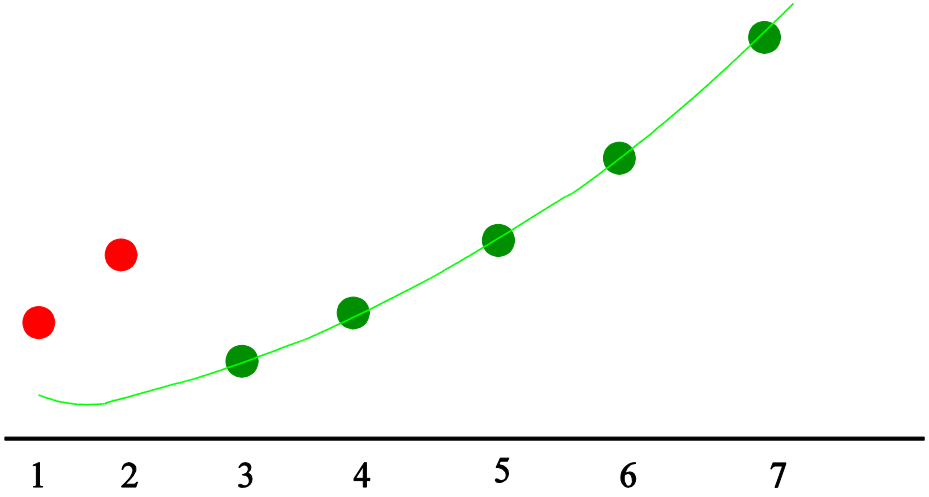
- Ein Polynom vom Grad < 3 ist durch seine Werte an drei Stellen eindeutig bestimmt.
- Stimmen zwei Polynome vom Grad < 3 an drei Stellen überein, so sind sie gleich.
- Für drei Stellen darf man die Werte beliebig vorgeben: Interpolationspolynom.
- Zwei verschiedene Polynome vom Grad < 3 schneiden sich höchstens zweimal.

Fehlerkorrigierende Kodes (Reed-Solomon)

- Ich will 1 1 3 senden.
- Bestimme das eindeutige Polynom vom Grad < 3 mit p(1) = 1, p(2) = 1, p(3) = 3.
- $p(x) = x^2 - 3x + 3$
- Sende 1 1 3 p(4) = 7 p(5) = 13 p(6) = 19, p(7) = 31.
- Bei der Übertragung passieren 2 Fehler. Der Empfänger erhält

4 7 3 7 13 19 31.

- Der Empfänger erhält 4 7 3 7 13 19 31
- Für jedes Tripel von Werten interpoliert er:
- p(3) = 0, p(5) = 13, p(7) = 31 → richtig. Polynom
- p(1) = 4, p(5) = 13, p(7) = 31 → falsches Polynom
- 10 mal kommt das richtige Polynom raus, 25 mal ein falsches. Aber jedes falsche höchstens 4 mal. Mehrheitsentscheid.



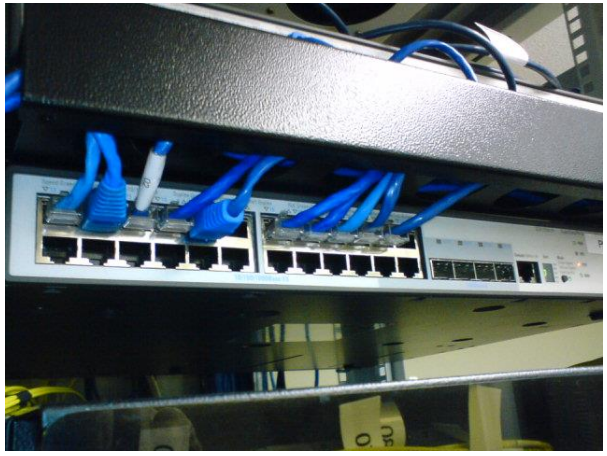
Auf einem falschen Polynom liegen höchstens 2 grüne Punkte, also insgesamt höchstens 4 Punkte.

Ein Geheimnis teilen

- Möchte n Personen ein Geheimnis geben, so dass es je k rekonstruieren können, aber k – 1 es nicht können.
- Sei g das Geheimnis. Wähle zufällige Zahlen a1 bis a(k-1) und bestimme das eindeutige Polynom p vom Grad < k mit p(0) = g und p(i) = ai für $1 \leq i \leq k - 1$.
- Gib der i-ten Person das Paar (i,p(i)), $1 \leq i \leq n$.

MAC (media access control) Adressen

- Im Ethernet hört jeder alles auf der Leitung
- Konfliktauflösung.
- Jedes Gerät hat eine eindeutige MAC Adresse (von Geburt an).
- Datenpakete haben einen Adresspräfix. Prozessor holt sich die für ihn bestimmten Nachrichten von der Leitung.



Internet Protocol (IP)

- Bietet Paket-Kommunikation *zwischen* Netzwerken
- Egal ob die Technik gleich ist oder nicht (Ethernet vs. WLAN).
- Best Effort, Keine Garantien:
  - Pakete gehen verloren
  - Pakete kommen doppelt an
  - Reihenfolge kann sich ändern

IP Adressen

- Wie Telefonnummern für Computer
- 32 Bits für die Adresse
  - Vier Zahlen zwischen 0 und 255
  - Zum Beispiel *139.19.14.56 = MPI-INF*
  - Regionales Clustering
  - Hat man nicht von Geburt an (MAC-Adresse), sondern bekommt man zugewiesen
- Ungefähr 4 Milliarden mögliche Adressen
- Bald aufgebraucht: Umstieg auf 128 Bits

IP Routing

- Jeder Router (Verteiler) hat eine Tabelle

Ziel	Link	Distanz
192.168.*.*	1	15
192.169.*.*	2	5
192.170.*.*	1	12

- Ist Ziel in meinem Netz? Direkt an MAC.
- Sonst in der Tabelle nachschlagen und auf entsprechendem Ausgabelink weiterleiten.

Routing Information Protocoll

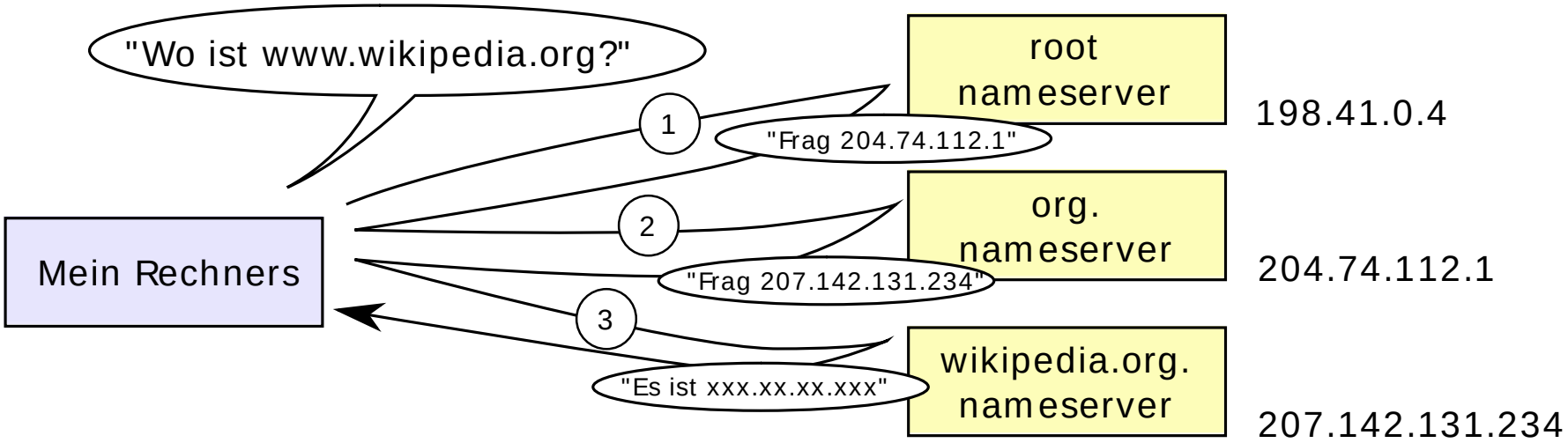
- Das Netz ändert sich ständig, z.B. Reparaturen oder neue Hardware.
- Router berechnen kontinuierlich kürzeste Pfade im Netz (kurz = wenige Hops).
- Alle 30 Sekunden: Tabelle an alle Nachbarn weiterreichen.
- Update: wenn mein Nachbar einen deutlich besseren Weg zu einem Ziel kennt, schicke ich die entsprechenden Pakete in Zukunft an ihn.

Transmission Control Protocol (TCP)

- Zuverlässige Datenübertragung zwischen Rechnern
  - Pakete nummerieren → Reihenfolge
  - Pakete mit Rückschein
  - Bleiben Bestätigungen aus → Neu senden

DNS

- Telefonbuch für IP Adressen
  - Übersetzt www.google.de in 173.194.35.151
- “Nameserver” speichern Tabellen
  - Tabelle enthält entweder Paar (Name,IP).
  - Oder Verweis auf Nameserver (mit .de gehst du besser zur Telekom).
  - Lokales Telefonbuch versus Auskunft.
- Jeder Computer hat eine Liste mit Nameservern.


"Wo ist www.wikipedia.org?"
Mein Rechners
1
"Frag 204.74.112.1"
root nameserver
198.41.0.4
2
"Frag 207.142.131.234"
org. nameserver
204.74.112.1
3
"Es ist xxx.xx.xx.xxx"
wikipedia.org. nameserver
207.142.131.234

Zwischenstand

- Ethernet und WLAN um im lokalen Netzwerk zu reden.
- IP um zwischen Netzwerken Pakete zu schicken.
- TCP um zuverlässig über IP zu reden.
- DNS um IP Adressen nachzuschlagen.

Email

- Post an mehlhorn@gmx.de schicken.
- Mailprogramm fragt Nameserver nach gmx.de und schickt die email an gmx.de.
- gmx speichert alle emails an mehlhorn in dessen Postfach.
- Ich hole sie von dort ab.

Hypertext Transfer Protocol, HTTP

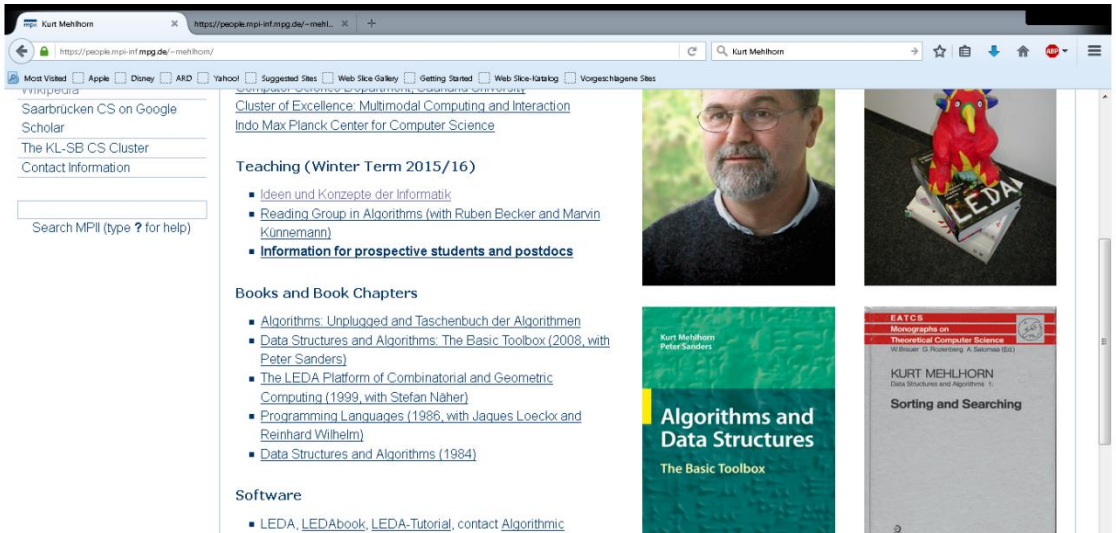
- HTTP ist ein Protokoll zur Übertragung von Daten auf der Anwendungsschicht über ein Rechnernetz.
- Es wird hauptsächlich eingesetzt, um Webseiten (Hypertext-Dokumente) aus dem World Wide Web (WWW) in einen Webbrowser zu laden.
- Webseiten sind in HTML kodiert.

Hypertext (HTML)

- “Sprache” in der Webseiten beschrieben sind.
- Der Text legt die Struktur der Webseite fest (Überschriften, Gliederung in Abschnitte, Tabellen, …) aber nur die ungefähre Darstellung.
- Webseiten enthalten Text, Bilder, Verweise, klickbare Objekte, …
- Browser berechnet Details der Dastellung, etwa Zeilenumbrüche, ….

Ausschnitt aus meiner Webseite

```
<H2><A>Books and Book Chapters</A></H2>
<UL type=circle>
  <li><a href="AlgorithmsUnplugged.html">Algorithms: Unplugged and Taschenbuch der Algorithmen </a></li>
   <li><a href="Toolbox.html">Data Structures and Algorithms: The Basic Toolbox (2008, with Peter Sanders) </a></li>
   <li><a href="LEDAbook.html">The LEDA Platform of Combinatorial and Geometric Computing (1999, with Stefan N&auml;her)</a></li>
  </ul>
```



Dynamische Elemente

- Mausbewegungen, Klicks etc. werden vom Betriebssystem verwaltet
- Browser wird über “Events” benachrichtigt
- Darstellung kann sich dynamisch ändern
  - Seite muss (effizient!) neu gezeichnet werden
- Klicken löst Aktionen aus
  - Zum Beispiel werden Videos abgespielt.

HTTPS versus HTTP

- http: unverschlüsselte Übertragung. Problematisch bei offenen WLANs
- S = secure
- Bietet
  - Authentifizierung der Partner.
  - Verschlüsselte Kommunikation.
- Empfehlung: HTTPS Everywhere benutzen.

Zusammenfassung