max planck institut
informatik

Ideen und Konzepte der Informatik

Kurt Mehlhorn

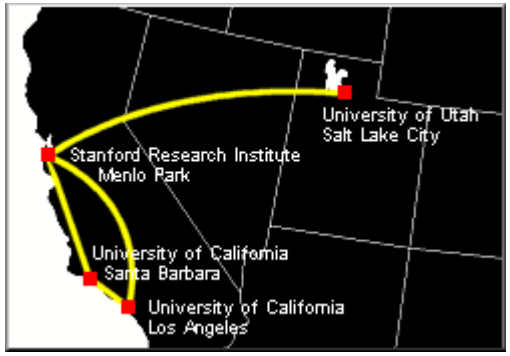
Internet und World Wide Web

Folien zum Teil von Kosta Panagiotou

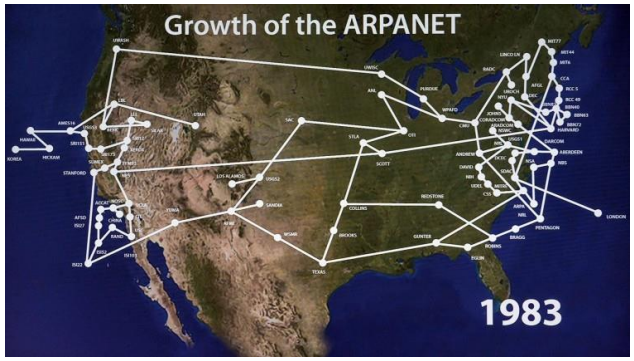
Das Internet und das World Wide Web

Internet = die Infrastruktur des digitalen Zeitalters, Datenautobahn + Datentransport, alles was es braucht, damit Inhalte bei Ihnen ankommen.

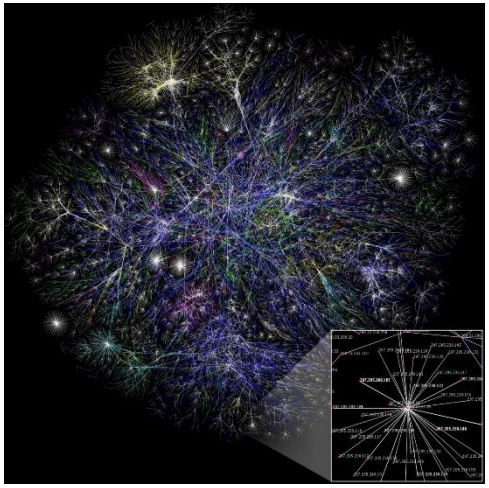
WWW = die Welt der Inhalte identifiziert durch URLs (universal resource locator)



Arpanet 1973



Arpanet 1983



Internet heute

Überblick

- Geschichte des Internets
- Datenübertragung im Internet
  - zwischen zwei Rechnern
  - zwischen Rechnern in einem Netzwerk
  - zwischen Netzen im Internet
  - und mit Garantien (Fehlerbehandlung, Fehlerkorrektur)
- Das World Wide Web
  - Aufbau von Webseiten
  - Darstellung im Webbrowser
  - E-Mail

Geschichte des Internets I

-- 1968: Großrechner in Hochschulen, Forschungszentren, großen Firmen, Militär, kaum Standardisierung, nur für Spezialisten.

1968 – 1983: Arpanet (Advanced Research Projects Agency Network), erstes Computer-Netzwerk, 4 US Universitäten.

- Packet Switching als neue Übertragungstechnik
- Standardisiertes Übertragungsprotokoll, Vinton Cerf, Turing Award
- Standardisierung bei Betriebssystemen und Programmiersprachen: Unix, C.
- Hauptanwendung: E-Mail
- Parallel: kommerzielle Netze, Börsen, Nachrichten-agenturen
- 73: KM spielt Schach auf einem Rechner, der 70 km entfernt ist.

Geschichte des Internets II

1981 – 1993:

- TCP/IP, Übertragungsprotokoll
- DNS, Domain name server, menschenlesbare Rechneradressen, mpi-inf.mpg.de statt 192.172.1.1
- Usenet, erste Webforen → soziale Netzwerke
- E-Mail setzt sich durch
- Erste Rechner für den Arbeitsplatz und für Privatpersonen
- Anschluss von Privatwohnungen ans Netz

Geschichte des World Wide Webs

1989 –

- Kommerzialisierung, Netze nicht mehr als Forschungsinfrastruktur sondern als Rückgrat der Wirtschaft
- Erfindung des WWW am CERN
- Tim Berners Lee, Turing Award
- Ziel: vereinfachter Datenaustausch zwischen Physikern an verschiedenen Orten.

- Hypertext, Seitenbeschreibungssprache HTML, das Transferprotokoll HTTP, die URL, der erste Browser WorldWideWeb (Darstellung von Inhalten) und der erste Webserver CERN httpd (Bereitstellung von Inhalten).
- 1993: erster graphikfähiger Browser

Die erste Webseite, CERN (1992)

World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists , Policy , November's W3 news , Frequently Asked Questions .

What's out there?
Pointers to the world's online information, subjects , W3 servers, etc.
Help
on the browser you are using
Software Products
A list of W3 project components and their current state. (e.g. Line Mode ,X11 Viola , NeXTStep , Servers , Tools , Mail robot , Library)
Technical
Details of protocols, formats, program internals etc
Bibliography
Paper documentation on W3 and references.
People
A list of some people involved in the project.
History
A summary of the history of the project.
How can I help ?
If you would like to support the web..
Getting code
Getting the code by anonymous FTP , etc.

Darstellung ist modern. Auf info.cern.ch kann man das Orginal im Orginalbrowser sehen.

Geschichte des WWW, II

Ab 2003

- Social Media Platformen, Facebook (seit 2004),
  Twitter (seit 2006), YouTube (seit 2005), Instagram (seit 2010)
- Suchmaschinen wie Google (seit 1997)
- Hochentwickelte Browser: Firefox, Chrome, Safari, Explorer
- Nutzergenerierte Inhalte
- Iphone (2007), weite Verbreitung von Smartphones, Android (seit 2008)
- mobile Breitbandsysteme und mobiles Internet (seit 2010)

Datenübertragung im Internet

- 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!
- Übertragungsfehler müssen repariert werden.
- Konstruktion von Datennetzen (allgemeiner (Informatik)-Systemen) geschieht in Schichten.

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 Installationsdienst für Häuser. Architekt nutzt Installationsdienst und bietet Bäder. Normen erleichtern die Zusammenarbeit.

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 zur Paketzustellung.
- Transport Layer
  - Fehlertolerante Datenübertragung.
- Data Layer
  - Kommunikationsprotokoll zwischen Browser und Server, Dienste für den Endnutzer.

Ethernet, ein populäres Netzwerk

- Kabelgebunden
- +5V = 1, – 5V = 0
- 1 Megabit – 100 Gigabits pro Sekunde

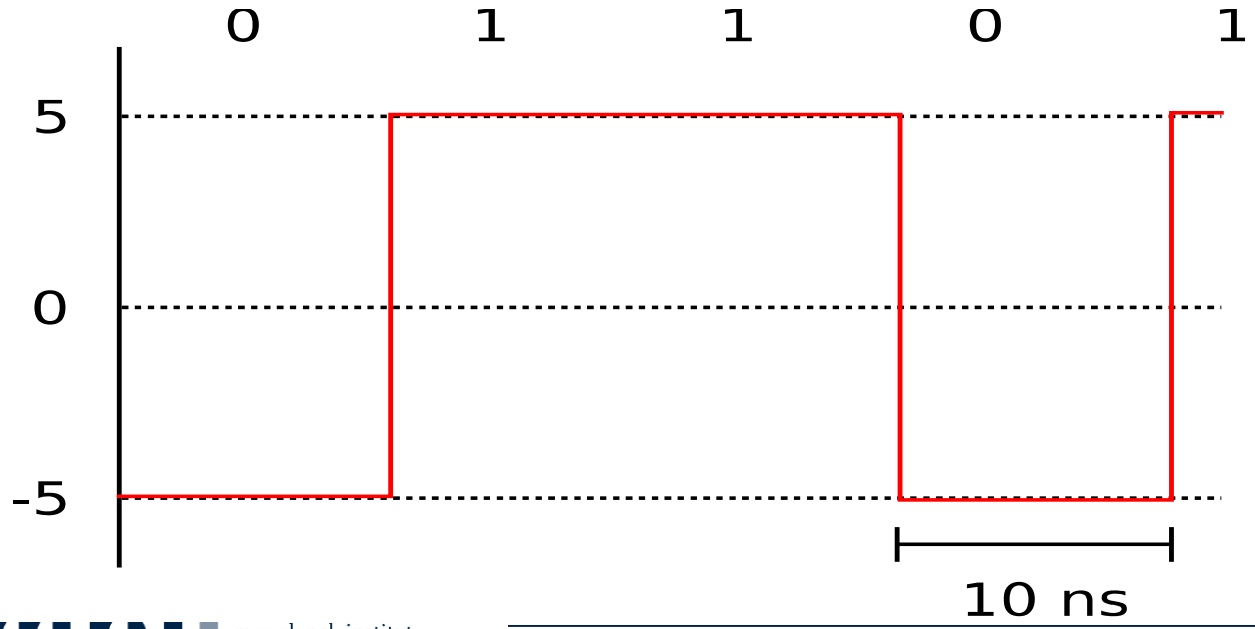


Abbildung ist stark idealisiert

Probleme

- Uhren:
  - Wann messe ich die Spannung?
  - Welche Uhrenqualität braucht man?
  - 1 000 000 Einsen = 10^{-2} Sekunden 5V,
    nicht 10^{-2} Sekunden + 10 ns
- Störungen
  - Sollte das eine 1 sein oder hat jemand den Föhn angemacht?

Selbstsynchronisierung
billige Uhren tun‘s 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; selbstsynchronisierend

010101110100110 →

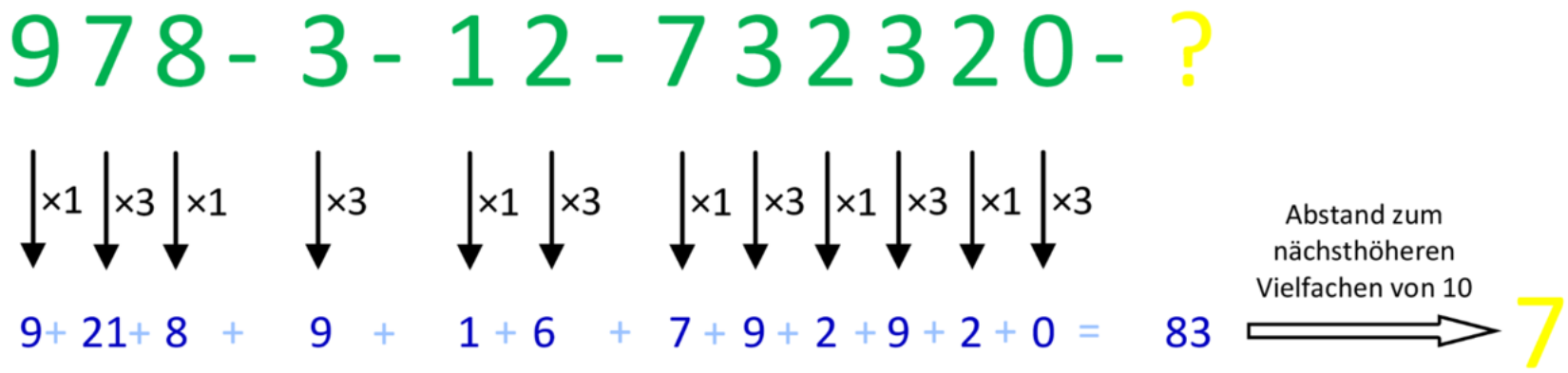
Störungen

- 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 Paket: Neuübertragung
- Oder fehlerkorrigierende Kodes; siehe übernächste Folie

Prüfsummen

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

Beispiel: Prüfziffer bei der ISBN-13



- IBAN ist ähnlich: DE02 590 200 90 000 8703 620

Fehlerkorrigierende Kodes (Reed-Solomon)

- Wir wollen eine Folge von *k* Zahlen senden (Zahlen statt Bits vereinfacht die Mathematik).
- Bis zu *d* Zahlen werden falsch übertragen, z.B. *d* = 5. Wir möchten die Nachricht immer noch lesen können.
- Erste Lösung: schicke jede Zahl 2*d* + 1 mal.
- Aufwendig: Für *d* = 5 wird die Nachricht 11 mal so lang.
- Reed-Solomon schickt nur 2*d* zusätzliche Zahlen. Ich zeige die Lösung für *k* = 2 und *d* = *2* und dann allgemein.

Mathematischer Hintergrund (k = 2)

- Eine Gerade ist durch zwei Punkte bestimmt.
- Durch zwei beliebige Punkte geht eine Gerade.
- Stimmen zwei Geraden an zwei Punkten überein, so sind sie gleich.
- Zwei verschiedene Gerade schneiden sich höchstens einmal.

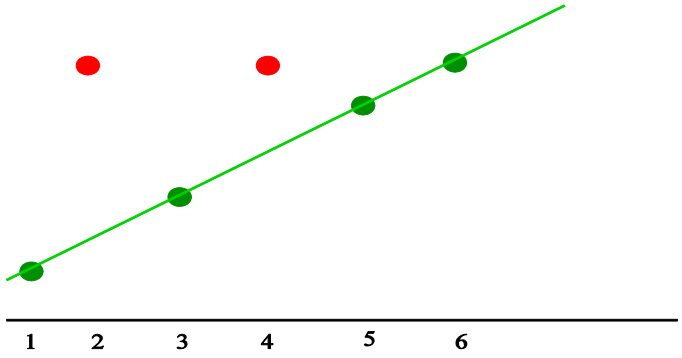
Fehlerkorrigierende Kodes (Reed-Solomon)

- Ich will 1 2 senden.
- Bestimme die eindeutige Gerade p mit $p(1) = 1$, $p(2) = 2$.
- $p(x) = x$.
- Sende 1 2 $p(3) = 3$, $p(4) = 4$, $p(5) = 5$, $p(6) = 6$.
- Bei der Übertragung passieren 2 Fehler. Der Empfänger erhält

  1 6 3 6 5 6

Fehlerkorrigierende Kodes (Reed-Solomon)

- Der Empfänger erhält 1 6 3 6 5 6. Für jedes Paar von Werten bestimmt er die Gerade. Es gibt 6 x 5 / 2 = 15 Paare.
- $p(1) = 1$, $p(3) = 3 \rightarrow$ richtige Gerade
- $p(1) = 1$, $p(4) = 6 \rightarrow$ falsche Gerade
- Auf der richtigen Gerade liegen 4 (grüne) Punkte. Auf einer falschen Gerade liegt höchstens ein grüner Punkt, also insgesamt höchstens 3 Punkte (zwei rote und ein grüner).



Also wird die richtige Gerade öfter gefunden als jede falsche.

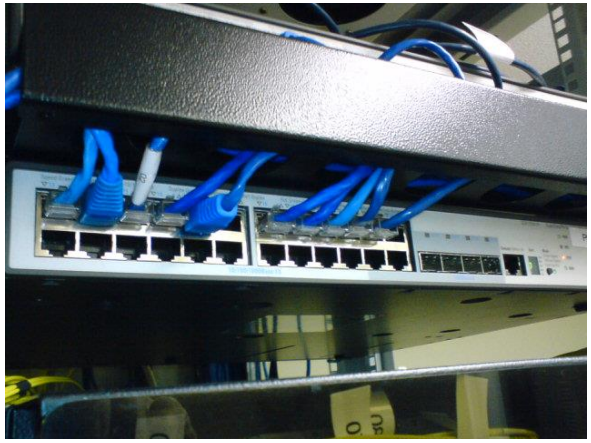
Mehrheitsentscheid

Ein Geheimnis teilen

- Weitere Anwendung des Prinzips: Möchten Bob und Alice ein Geheimnis mitteilen, so dass es einer allein nicht rekonstruieren kann (Tresor mit 2 Schlüsseln).
- Sei *g* das Geheimnis. Wähle eine zufällige Zahl a und gib Bob die Zahl g – a und Alice die Zahl g + a.
- Zusammen können sie g bestimmen, da (g – a + g + a)/2 = g.
- Einer allein weiß gar nichts: g + a ist eine zufällige Zahl.

MAC (media access control) Adressen

- Im Ethernet und im WLAN hört jeder alles.
- Konfliktauflösung, ALOHA Protokoll
- Jedes Gerät hat eine eindeutige MAC Adresse (von Geburt an).
- Datenpakete haben feste Länge und ein Adresspräfix. Prozessor holt sich die für ihn bestimmten Nachrichten von der Leitung.
- Inhalte sollten verschlüsselt sein.



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

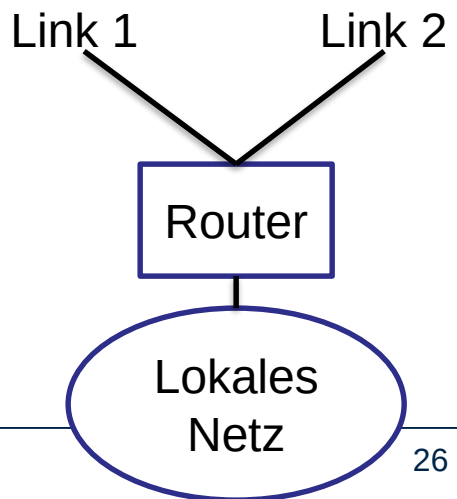
- Wie Telefonnummern für Computer
- 32 Bits für die Adresse
  - Vier Zahlen zwischen 0 und 255
  - Zum Beispiel *139.19.14.56 = MPI-INF*
  - Regionales Clustering
  - Hat man nicht von Geburt an (wie bei der MAC-Adresse), sondern bekommt man zugewiesen.
- 32 Bit → ungefähr 4 Milliarden mögliche Adressen; daher Übergang zu 128 Bit.

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 Protocol

- 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. Wenn sich mein Nachbar 60 Sekunden nicht meldet, schicke ich nichts mehr an ihn.

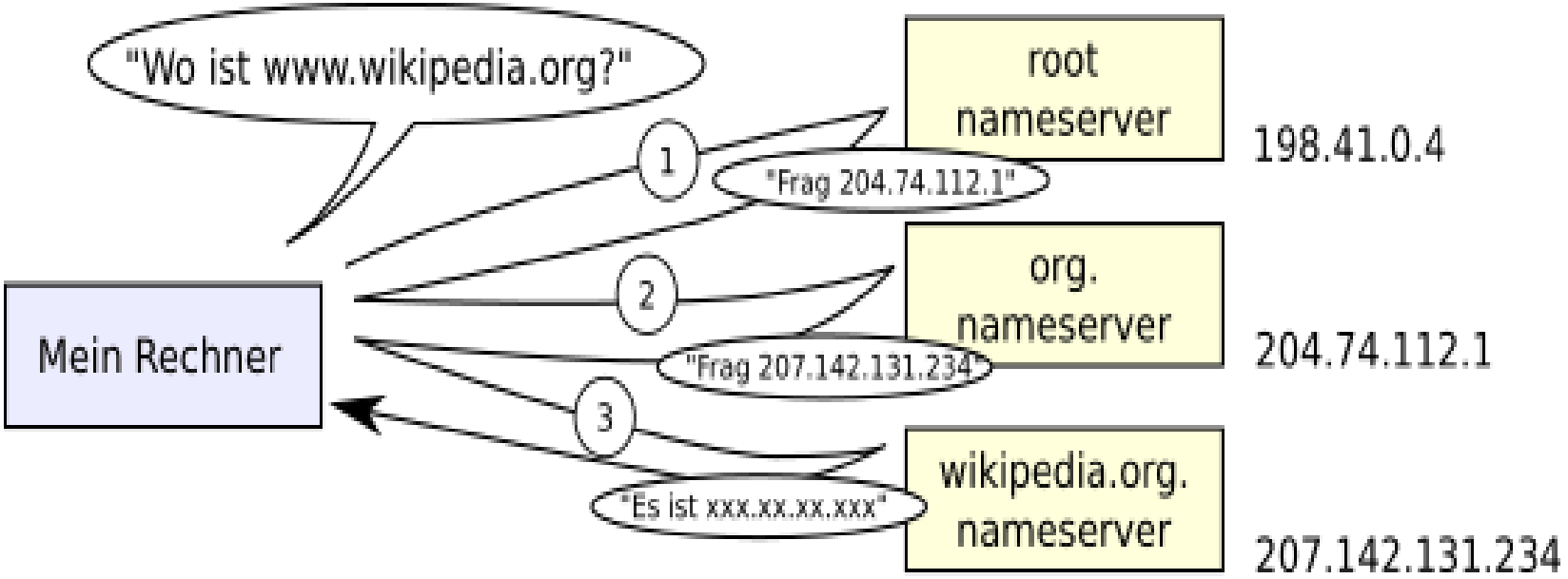
Transmission Control Protocol (TCP)

- 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.

Nachschlagen von Wikipedia.org



Man geht zuerst zum Root-Nameserver. Der verweist einen weiter.

Zusammenfassung Internet

- 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.

- Und nun zu den Diensten: E-Mail, World Wide Web.

E-Mail

- Post an *mehlhorn@mpi-inf.mpg.de* oder *mehlhorn@gmail.com* schicken.
- Mailprogramm fragt Nameserver nach mpi-inf.*mpg.de* und schickt die E-Mail an den Mailserver von mpi-inf.mpg.*de*. (analog gmail)
- Mailserver von mpi-inf.mpg.de speichert alle E-Mails an mehlhorn in dessen Postfach.
- KM holt sie dort ab oder KM liest seine Mail auf dem Server (bei gmail).

Hypertext Transfer Protocol, HTTP

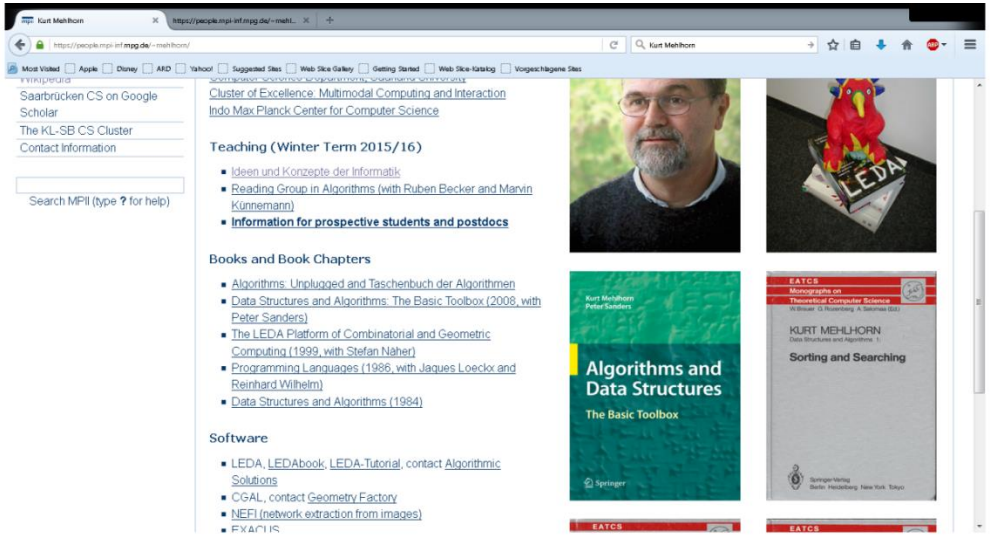
- HTTP ist ein Protokoll zur Übertragung von Daten auf der Anwendungsschicht über ein Rechnernetz.
- Es wird hauptsächlich eingesetzt, um Webseiten (Hypertext-Dokumente) aus dem World Wide Web (WWW) in einen Webbrowser zu laden.
- Webseiten sind in HTML (Hypertext Markup Language) kodiert.
- Hypertext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

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 Darstellung, etwa Zeilenumbrüche, ….

Ausschnitt aus KMs 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>
```



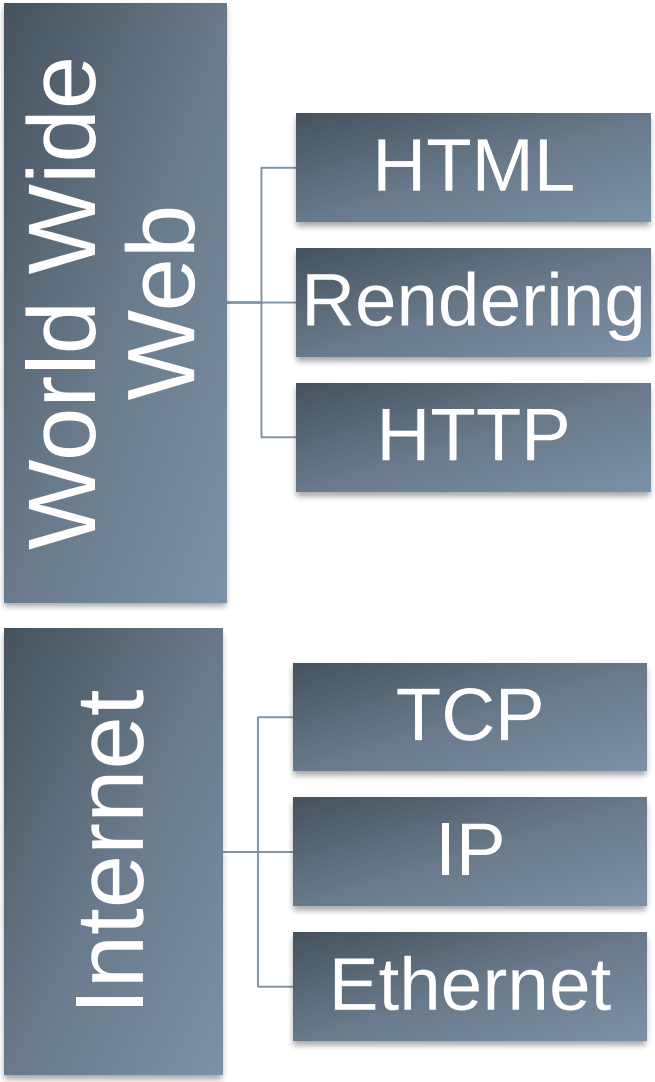
Dynamische Elemente

- 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. Kommt aus der Mode.
- S = secure
- Bietet
  - Authentifizierung der Partner
  - Verschlüsselte Kommunikation
- Empfehlung: HTTPS Everywhere benutzen. Viele Browser verweigern das Gespräch mit http-Seiten.

Zusammenfassung



Mathematischer Hintergrund (k = 3)

- 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.

Mathematischer Hintergrund (k = 3)

- Ein Polynom vom Grad < 3 ist durch seine Werte an drei Stellen eindeutig bestimmt.
- $p(x) = a_2x^2 + a_1x + a_0$, Polynom vom Grad 2; a_2, a_1, a_0 sind die Koeffizienten.
- Wert an der Stelle 5: $p(5) = 25a_2 + 5a_1 + a_0$.
- Falls $p(0) = 2, p(2) = 16, p(-1) = 4$, dann $a_2 = 3, a_1 = 1, a_0 = 2$.

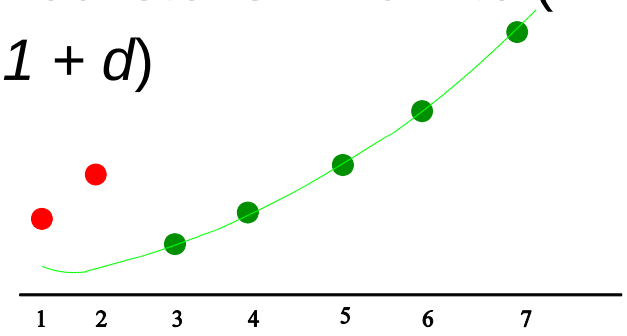
Fehlerkorrigierende Kodes (Reed-Solomon)

- 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 $p(4) = 7$, $p(5) = 13$, $p(6) = 21$, $p(7) = 31$.
- Bei der Übertragung passieren 2 Fehler. Der Empfänger erhält

  4 7 3 7 13 21 31.

Fehlerkorrigierende Kodes (Reed-Solomon)

- Der Empfänger erhält 4 7 3 7 13 19 31. Für jedes Tripel von Werten interpoliert er. Es gibt 35 Tripel.
- $p(3) = 3$, $p(5) = 13$, $p(7) = 31 \rightarrow$ richtiges Polynom
- $p(1) = 4$, $p(5) = 13$, $p(7) = 31 \rightarrow$ falsches Polynom
- Auf dem richtigen Polynom liegen mindestens 5 Punkte (mindestens $k + d$). Auf einem falschen Polynom liegen höchstens 4 Punkte (zwei rote und zwei grüne, allgemein $k - 1 + d$)



Daher wird das richtige Polynom öfter gefunden als jedes falsche.

Mehrheitsentscheid.

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 a_1 bis a_{k-1} und bestimme das eindeutige Polynom p vom Grad $< k$ mit $p(0) = g$ und $p(i) = a_i$ für $1 \leq i \leq k - 1$.
- Gib der i-ten Person das Paar $(i, p(i))$, $1 \leq i \leq n$.
- Anwendung: g ist ein Schlüssel. Je k Teilnehmer können schließen, aber keine $k - 1$ können es.