

Verteilte Systeme

Distributed Systems

Prof. Dr. Georg Hinkel
12.04.2023

ÜBER MICH

- Studium und Promotion am Karlsruher Institut für Technologie
 - Dissertation: Implizit inkrementelle Modellanalysen und –transformationen
 - Mitarbeit im Human Brain Project (Neurorobotik) über das Forschungszentrum Informatik (FZI)
- Industrieerfahrung bei Tecan Software Competence Center GmbH
 - 2018-2022
 - Software-Technologieentwicklung, Entwicklung von Konzepten und Prototypen im Bereich Laborautomatisierung
 - Teil der Arbeitsgruppe für SiLA2-Standard
- Seit 2023: Professur für Angewandtes Software Engineering



ADMINISTRATIVES

Verteilte Systeme

Prof. Dr. Georg Hinkel
12.04.2023

- Pflichtveranstaltung Bachelor
 - Angewandte Informatik
 - Informatik Technischer Systeme
- Prüfung
 - Präsenzklausur (benotet, 90min)
 - Praktikum (benotet)
- Inhaltliche Voraussetzungen
 - Rechnernetze und Telekommunikation
 - Programmierung: bspw. Java für das Praktikum

Warnung

Die Vorlesung wurde in Hinblick auf die kommende PO 2024 inhaltlich angepasst. Inhalte unterscheiden sich z.T. von der Veranstaltung von Prof. Kaiser

- Vorlesung
 - 2 SWS
 - Fr 8:15 – 9:45
 - Folien über Stud.IP verfügbar
- Praktikum
 - 2 SWS
 - Theorieübungen und praktische Aufgaben
 - Praktikumsnote basierend auf praktischen Aufgaben
 - Praktische Übungen in C, Java oder C#
 - Übungsblätter über Stud.IP
- Aufwand 5 CP ~ 150h
 - 42h Anwesenheit
 - ~108h Selbststudium inkl. Prüfungsvorbereitung

- Präsenzklausur
 - 90 min
 - 50% der Punkte garantieren Bestehen
- Praktikum
 - Bewertung von praktischen Aufgaben mit Punkten
 - Präsentation von theoretischen Aufgaben
 - 75% Anwesenheitspflicht
 - 50% der Punkte auf alle Praktikumsaufgaben garantieren Bestehen
- Gesamtnote
 - 60% Klausur
 - 40% Praktikum

- Folien
 - Verfügbar über Stud.IP
- Übungsblätter
 - Verfügbar über Stud.IP
- Lehrbücher
 - Tanenbaum, van Steen: “Verteilte Systeme - Grundlagen und Paradigmen”, Pearson Studium, 2. Auflage, 2007, ISBN 978-3-8273-7293-2, 49,95€
 - Coulouris, Dollimore, Kindberg, Blair: “Distributed Systems - Concepts and Design”, Pearson Studium, 5. Auflage, 2012, ISBN 978-0132143011, 147,95€
 - HTTP/3.0 Explained, <https://http3-explained.haxx.se/de>
 - Michael Nygard: Release It!: Design and Deploy Production-Ready Software, O'Reilly Media, 2018, ISBN-13: 978-1680502398



- Kurs-Unterlagen bei Particify
 - <https://arsnova.hs-rm.de/p/24881251>
 - Code: 2488 1251
 - Q&A
 - Feedback zur Veranstaltung
 - Frageserien



EINFÜHRUNG

Verteilte Systeme

Prof. Dr. Georg Hinkel
19.04.2024

Agenda

- Geschichtliche Entwicklung
 - Vom ARPAnet zum Internet
 - Internet of Things
- Grundbegriffe Verteilter Systeme
 - Verteiltes Programm, verteilter Zustand
 - Transparenzarten nach ISO
- Standardisierung

Lernziele

- Grundbegriffe kennen und zuordnen können
- Historische Entstehung des Internets in groben Zügen wiedergeben können
- Standardisierungsprozesse erklären können

HISTORIE

Halbleitertechnologie: Leistung und Kosten

- Speicherchips
 - 1973: 4 kBit
 - 1985: 64 kBit
 - 1998: 64 MBit
 - 2008: 16 GBit
 - 2018: 128 GBit
- Gesetz von Moore (1965): Alle anderthalb Jahre verdoppelt sich die Zahl der Transistorfunktionen auf der gleichen Grundfläche
- Entwicklung der Kosten je Transistorfunktion auf ca. 1/10 alle vier Jahre
- Immer wieder Ende des Gesetzes vorausgesagt
- Neuere Technologien: Z-RAM, MRAM, FeRAM, ...

Entwicklung der CPU-Komplexität

S Our World
in Data

S Our World
in Data

S Our World
in Data



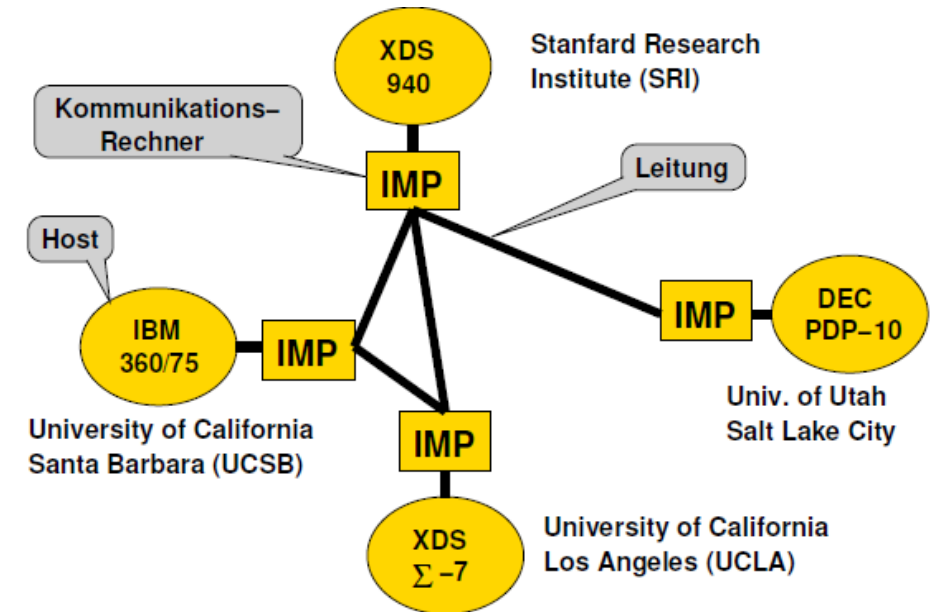
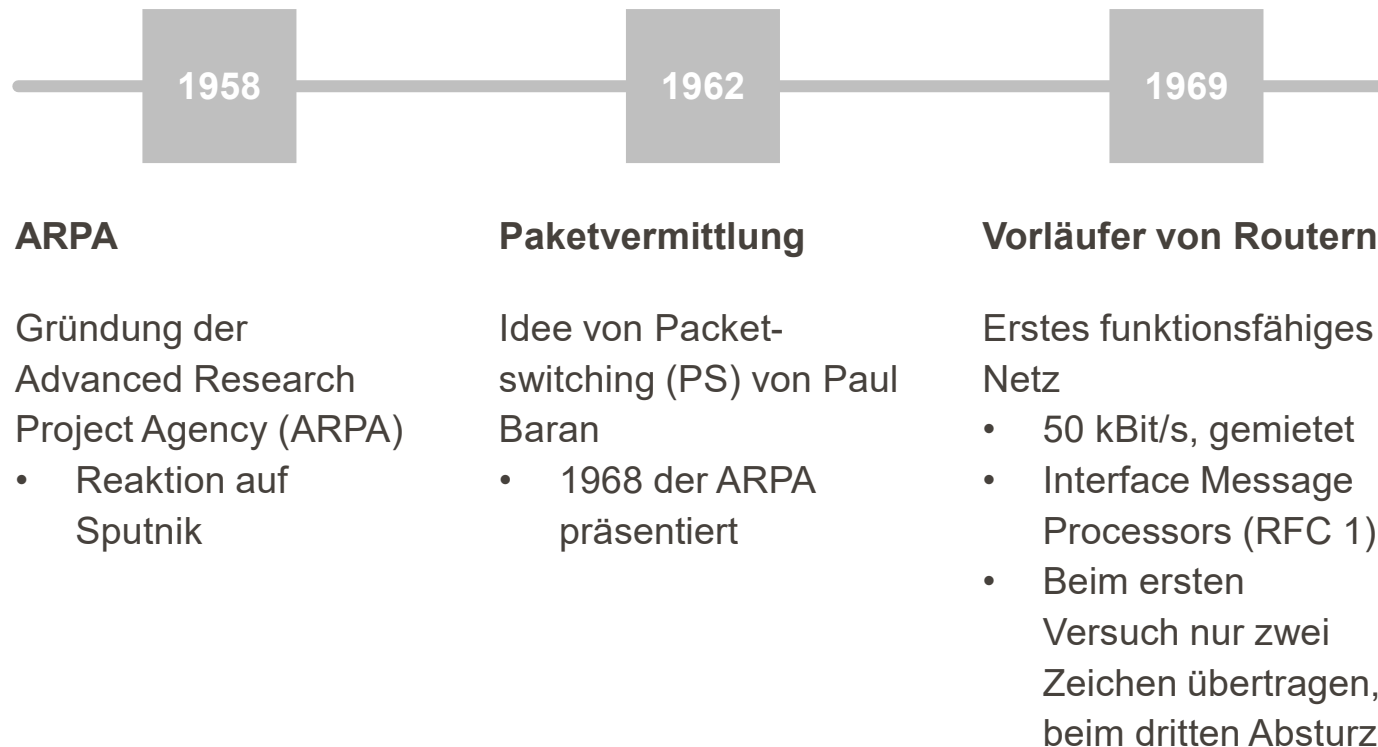
S Our World
in Data

S Our World
in Data

S Our World
in Data

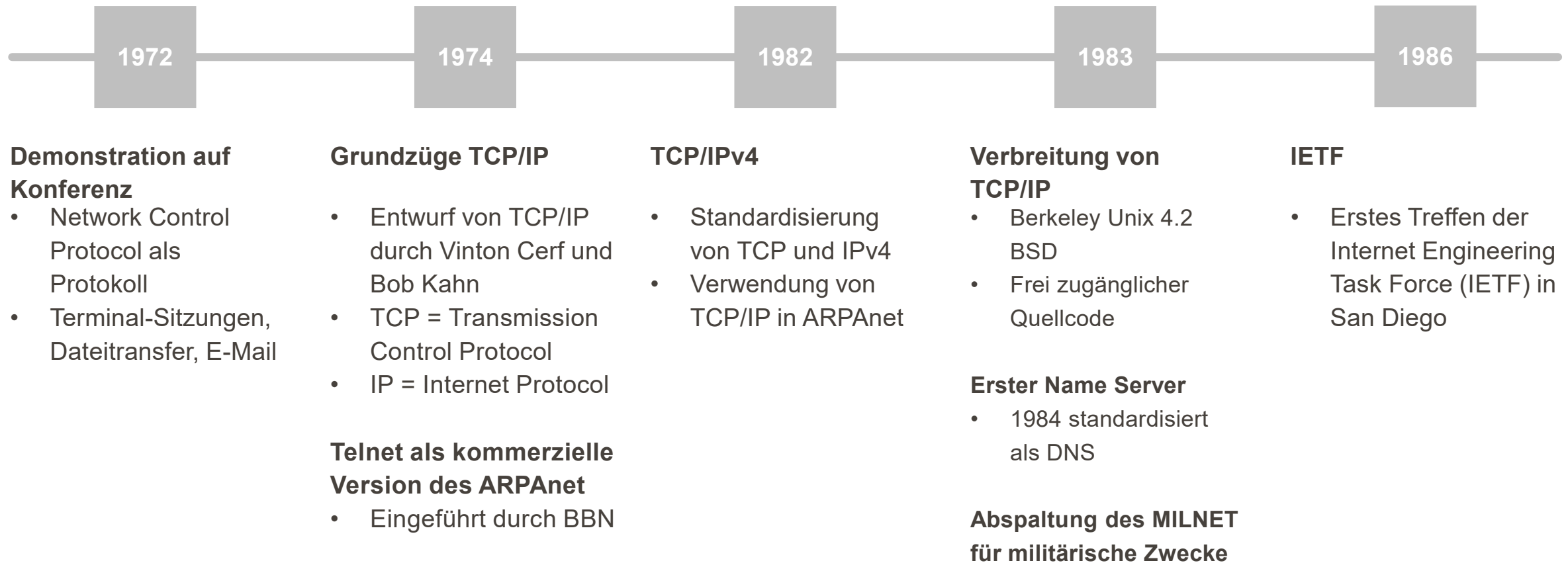
HISTORIE

Vom ARPANET zum Internet (I)



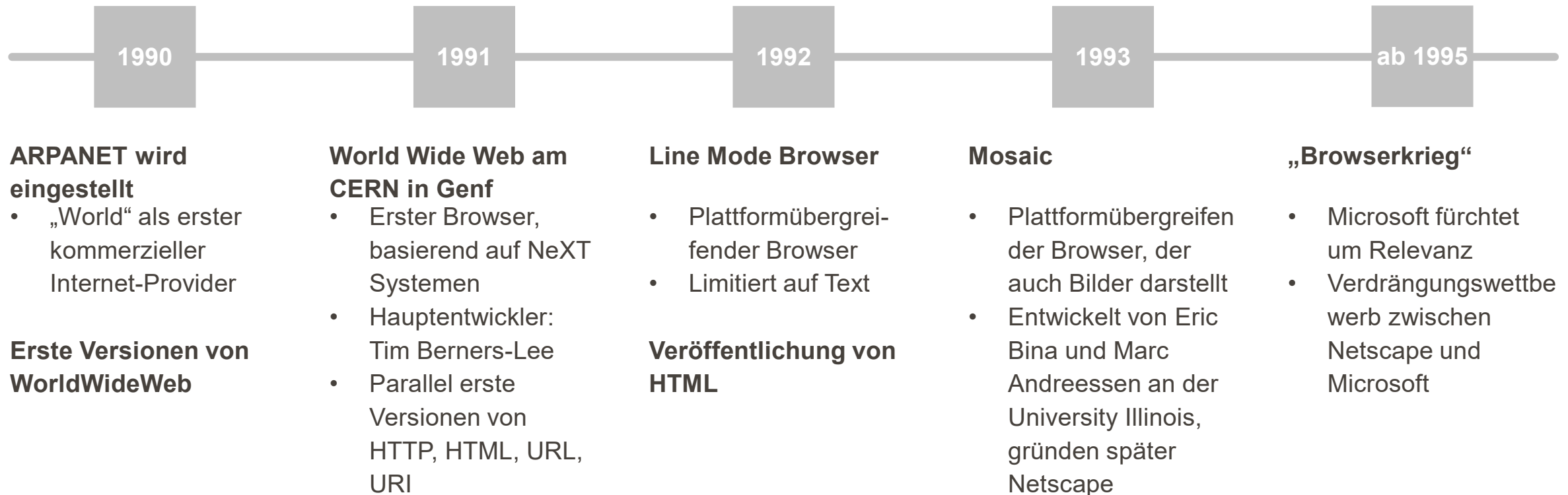
HISTORIE

Vom ARPANET zum Internet (II)



WAS WAR NOCHMAL DER UNTERSCHIED ZWISCHEN TCP UND UDP?

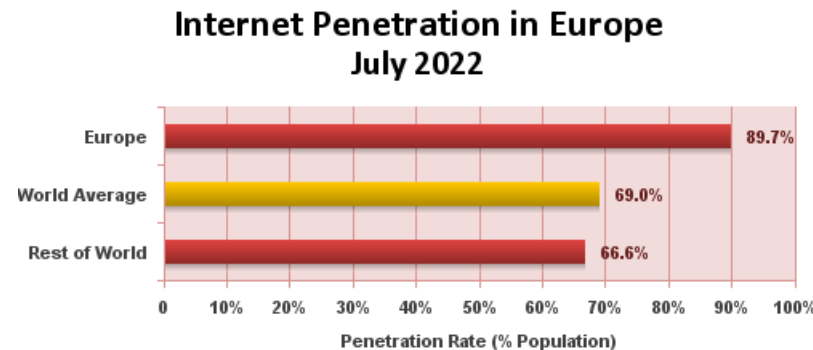




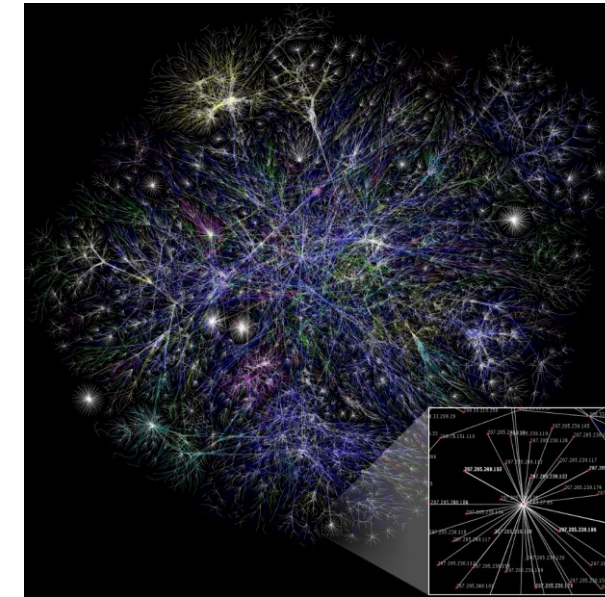
- Ursprung: Microsoft sieht eigene Relevanz bedroht
 - Geschäftsmodell von Microsoft Mitte der 90er: Betriebssystemlizenzen
 - Wenn Anwendungen nur noch als Webanwendungen realisiert werden, ist Betriebssystem egal
 - Entwicklung eines eigenen Browsers, um Netscape Navigator zu verdrängen
 - Massiver Innovationsschub
 - Erster Browserkrieg endet mit Niederlage von Netscape
- Umfangreiche Klagen gegen Microsoft
 - Bündelung Internet Explorer mit Windows
- „Zweiter Browserkrieg“
 - Alternative Browser (neben Internet Explorer) gewinnen Anfang 2000er Marktanteile
 - Standardisierung

WACHSTUM DES INTERNETS

- Autonome Systeme
 - Steter Zuwachs
 - Stabiler Kern, Wachstum an der Peripherie
- Internet-Nutzer
 - Mittlerweile mehr als zwei Drittel der Menschheit online
 - In Europa fast 90%



Source: Internet World Stats - www.internetworldstats.com/stats4.htm
Based on 5,475,899,417 world Internet users on July 31, 2022
Copyright © 2022, Miniwatts Marketing Group



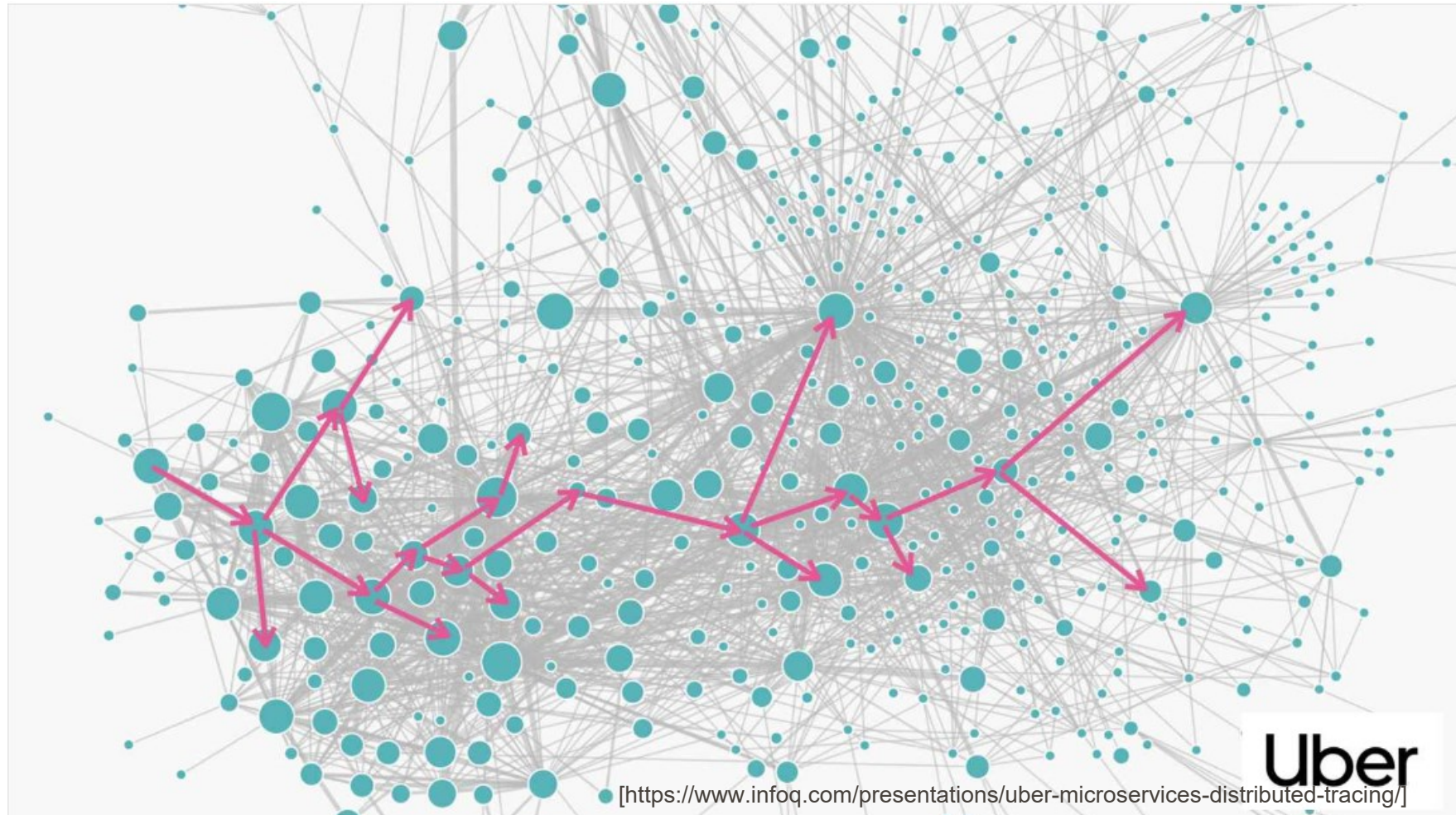
[Map of the internet, Barett Lyon, 2006]

- **1968** Richard („Dick“) Morley entwickelt Programmable Logic Controller (PLC) für Industriefertigungsanlagen
- **1982** An der Carnegie Mellon University wird ein Getränkeautomat mit dem Internet verbunden
- **1994** Gründung der OPC Foundation → verteilte Systeme für Automatisierungstechnik
- **1995** Veröffentlichung der ersten IPv6 Spezifikation
- **1996** Hewlett-Packard und Nokia veröffentlichen mit dem OmniGo 700LX und dem 9000 Communicator erste Smartphone-Vorläufer
- **1997** Kristofer S. J. Pister, Joe Kahn und Bernhard Boser präsentieren Forschungsantrag zu Smart Dust
- **1999** Kevin Ashton prägt den Begriff des Internet of Things (IoT)
- **2003** Walmart setzt RFID Chips für die Inventarisierung ein
- **2006** Veröffentlichung von OPC UA
- **2012** General Electric bringt den Begriff Industrial Internet of Things (IIoT) in Umlauf
- **2015** Börsengang von FitBit

- Personal Computer (PC, Laptop), Workstations
- Server, Großrechner (Mainframes)
 - Hochverlässliche Verarbeitung von Massendaten
 - Hoch- bis Höchstleistungs-Ein-/Ausgabe-Einheiten
 - Erbringen Dienstleistungsfunktionen in Rechnernetzen
 - Mainframes z.T. immer noch wegen Altprogrammen erforderlich (Legacy-Systeme)
- Supercomputer
 - Vielzahl von Prozessoren/Knoten
 - Hohe Verarbeitungsleistung
 - Beispiel: Wettervorhersage, Eiweißsimulationen
- Embedded Systems
 - Teil von Maschinen, Geräten, Anlagen
 - Typischerweise eingeschränkte Rechenleistung
 - Cyber-Physical Systems / Industrie 4.0

Heute zumeist agierend als Bestandteile verteilter Systeme

MICROSERVICES BEI UBER



WAS GENAU IST DENN NUN EIN VERTEILTES SYSTEM?

Definitionsversuche

*» A DISTRIBUTED SYSTEM IS A COLLECTION
OF AUTONOMOUS COMPUTING
ELEMENTS THAT APPEARS TO ITS USERS AS
A SINGLE COHERENT SYSTEM.«*

Maarten van Stehen, Andrew S. Tanenbaum

*» A DISTRIBUTED SYSTEM IS ONE IN WHICH
THE FAILURE OF A COMPUTER YOU
DIDN'T EVEN KNOW EXISTED CAN RENDER
YOUR OWN COMPUTER UNUSABLE.«*

Leslie Lamport

WARUM VERTEILTE SYSTEME?

Verteilung ist notwendig...

- Reichweite
 - 69% der Menschheit nutzt regelmäßig das Internet
- Ressourcen
 - Viele Aufgaben auf einer einzelnen Maschine nicht lösbar
- Ausfallsicherheit
 - Verteiltes System kann Ausfall einer Maschine verkraften
- Unabhängigkeit
 - Lose Kopplung erlaubt unabhängige Entwicklung
 - Verschiedene Programmiersprachen für verschiedene Zwecke / Teams

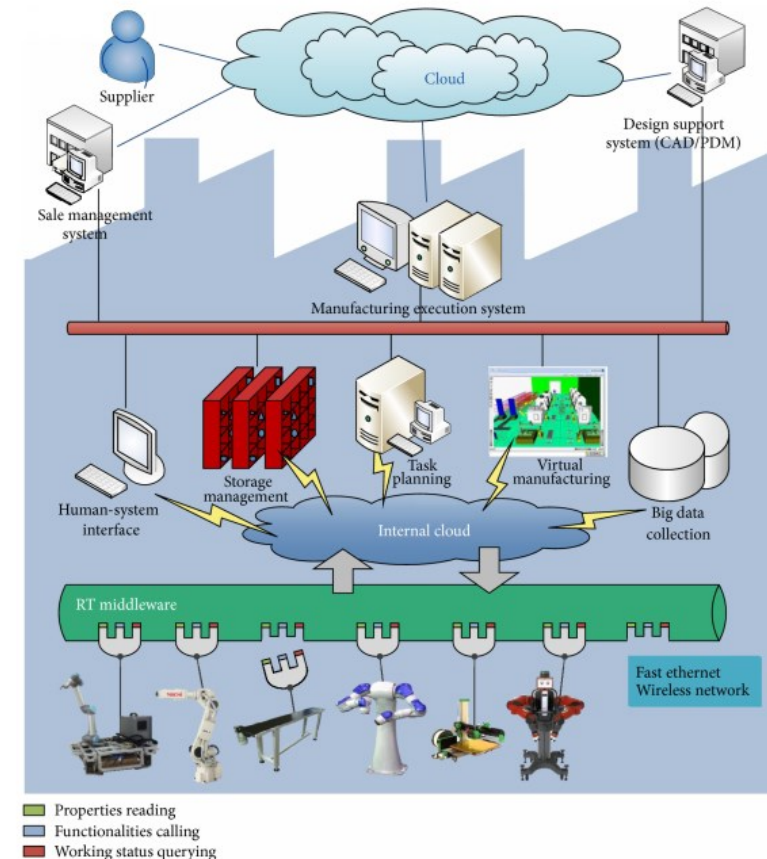
WARUM VERTEILTE SYSTEME?

...aber wir wollen sie eigentlich nicht sehen

- Transparenz → Unsichtbarkeit von Eigenschaften
 - **Ortstransparenz:** Keine Kenntnis des Ortes notwendig, Ressource kann mit Namen verwendet werden
 - **Zugriffstransparenz:** Form des Zugriffs ist unabhängig ob Komponente lokal oder entfernt
 - **Fehlertransparenz:** Eingetretener Fehler wird nicht sichtbar, sondern durch Redundanz maskiert
 - **Parallelitätstransparenz:** Nebenläufige Zugriffe teilen sich Ressourcen ohne sich gegenseitig zu stören
- Weitere Transparenzarten nach ISO: u.a. Migrationstransparenz, Replikationstransparenz, Nebenläufigkeitstransparenz, Skalierungstransparenz, Leistungstransparenz

BEISPIELE VERTEILTER SYSTEME

- Web
 - Geschäftsanwendungen
 - Kollaborative Systeme
 - Soziale Medien
 - E-Commerce
 - ...
- Internet of Things (IoT)
 - Wearables
 - Vernetzte Sensorik
 - Industrie 4.0
 - ...



[<https://commons.wikimedia.org/wiki/File:System-architecture-of-the-smart-factory.jpg>]

- Enge Kopplung
 - Softwarekomponenten, die durch gemeinsame Nutzung von Betriebsmitteln kommunizieren
 - Gemeinsam genutzte Objekte
 - Gemeinsam genutzter Speicher
 - Typischerweise im selben Prozess
- Lose Kopplung
 - Softwarekomponenten, die durch Nachrichtenaustausch (Message Passing) kommunizieren
 - Dadurch höhere Autonomie
- Verteiltes Programm / Verteilte Anwendung
 - Menge von lose gekoppelten Softwarekomponenten, die für die Lösung eines Problems zusammenarbeiten
 - Beinhaltet verteilten Zustand, verteilte Kontrolle bzw. Koordination
- Verteiltes System
 - Rechnernetz, was verteiltes Programm ausführt

- Viele Freiheitsgrade bei der Kommunikation in verteilten Systemen → Standardisierung notwendig
 - Übereinkunft zur Vereinheitlichung von Dokumenten, Verfahren, Protokollen, usw.
 - De-jure (Norm) oder de-facto („Industriestandard“)
 - Unterschiedlicher Gültigkeitsbereich



[Bild: Wikipedia, CC-SA, KMJ]

WARUM SIND STANDARDS WICHTIG?

- Kompatibilität / Interoperabilität
 - Zwischen Programmiersprachen / Software-Plattformen
 - Erschwert Vendor-Lock-in
- Kostensenkung
 - Wiederverwendbarkeit (Implementierungen, Dienste, Werkzeuge, ...)
 - Kürzere Einarbeitungszeit → kürzere Entwicklungszeit
 - Dafür häufig Einstiegshürde (Ausnahme: offene Standards)
- Höhere Qualität
 - Typischerweise umfangreiche Review-Zyklen

WIE ENTSTEHT EIN STANDARD?

- Gründung eines Gremiums / Arbeitsgruppe
 - Öffentlich oder privat, teilweise kostenpflichtig
- Abgrenzung des Standardisierungsgegenstandes
- Iterativer Prozess
 - Begutachtung / Revision
 - Veröffentlichung (frei zugänglich oder eingeschränkt)
 - Aktualisierung
 - Unterschiedlich formaler Prozess
 - Beispiel: RFC (Request for Comments), von akademischer Demut geprägt
- Gegebenenfalls Übertrag des Standards an anderes Standardisierungsgremium
 - Sichtbarkeit

- ISO (International Organization for Standardization)
 - Gegründet 1947, 164 Nationen
 - Alle Bereiche (weltweit)
- ITU (International Telecommunication Union)
 - Gegründet 1865, 196 Nationen
 - Technische Aspekte der Telekommunikation (weltweit)
- IEC (International Electrotechnical Commission)
 - Gegründet 1906, 80 Nationen
 - Elektrotechnik im weitesten Sinn, viele Normen zusammen mit ISO
- DIN (Deutsches Institut für Normung)
 - Gegründet 1917, über 2700 Mitglieder
 - Alle Bereiche (deutschlandweit)

RELEVANTE STANDARDISIERUNGSORGANISATIONEN

- IETF (Internet Engineering Task Force)
 - „Above the wire and below the application“
 - Freiwilligenvereinigung
- ICANN (Internet Corporation for Assigned Names and Numbers), inkl. IANA
 - Koordination und Vergabe von Adressen, Protokollnummern, Namen, etc.
 - Non-Profit-Organisation, bis 2016 der US-Regierung unterstellt
 - Vergibt auch Port-Nummern, Felder für Zertifikate, etc.
- IEEE (Institute of Electrical and Electronics Engineers)
 - Elektrotechnik und Informationstechnik
 - Weltweiter Berufsverband der Ingenieure (über 400.000 Mitglieder aus 160 Nationen)
 - Neben Standardisierung auch bspw. Verlag für wissenschaftliche Zeitschriften

RELEVANTE STANDARDISIERUNGSORGANISATIONEN

- OASIS (Organization for the Advancement of Structured Information Standards)
 - Dokumentenformate und Protokolle in der Telekommunikationstechnik
 - Non-Profit-Organisation
 - Wichtigste Standards: OpenDocument, BPEL, MQTT
- W3C (World Wide Web Consortium)
 - Webtechnologie
 - Industriekonsortium
- OMG (Object Management Group)
 - Systemübergreifende Objektorientierte Programmierung
 - Industriekonsortium
 - Wichtigste Standards: UML, CORBA, MDA

WORUM GEHT ES IN DER VORLESUNG?

- Wie programmiert man verteilte Systeme?
 - Kommunikationsmuster, Sockets, Architektur
- Wie arbeitet das World Wide Web?
 - HTTP/1.1, HTTP/2.0, QUIC, HTTP/3.0
- Wie arbeitet das Internet der Dinge?
 - MQTT, OPC UA, u.a.
- Wie kriegen wir das alles sicher hin?
 - Verschlüsselung, X.509 Zertifikate, TLS
- Wie erreichen wir Ortstransparenz?
 - Namen- und Verzeichnisdienste
- Wie erreichen wir Fehlertransparenz?
 - Resilience Patterns, Datenversionierung, verteilte Transaktionen
- Wie erreichen wir Parallelitätstransparenz?
 - Multiplexing
- Wie erreichen wir Zugriffstransparenz?
 - Remote Procedure Calls, gRPC

GLIEDERUNG

Datum	Vorlesung	Übungsblatt	Abgabe
19.04.2024	Einführung	HamsterLib	06.05.2024
26.04.2024	Netzwerkprogrammierung	Theorie	
03.05.2024	World Wide Web	HamsterRPC 1	20.05.2024
10.05.2024	Remote Procedure Calls	Theorie	
17.05.2024	Webservices	HamsterRPC 2	03.06.2024
24.05.2024	Fehlertolerante Systeme	Theorie	
31.05.2024	Transportsicherheit	HamsterREST	17.06.2024
07.06.2024	Architekturen für Verteilte Systeme	Theorie	
14.06.2024	Internet der Dinge	HamsterIoT	01.07.2024
21.06.2024	Namen- und Verzeichnisdienste	Theorie	
28.06.2024	Authentifikation im Web	HamsterAuth	15.07.2024
05.07.2024	Infrastruktur für Verteilte Systeme	Theorie	
12.07.2024	Wrap-Up	HamsterCluster (Bonus)	16.08.2024

- In welchem Jahrzehnt wurde TCP/IP standardisiert?
- In welchem Jahrzehnt wurde HTTP standardisiert?
- Was versteht man unter den Begriffen verteiltes System, verteilte Anwendung, enger Kopplung, und loser Kopplung?
- Warum werden Anwendungen häufig als verteilte Systeme implementiert?
- Erläutern Sie Ortstransparenz, Zugriffstransparenz, Fehlertransparenz und Nebenläufigkeitstransparenz!
- Erläutern Sie die Notwendigkeit von Standards!