Karlsruhe Institute of Technology

aifb

CRITICAL INFORMATION INFRASTRUCTURES RESEARCHGROUP

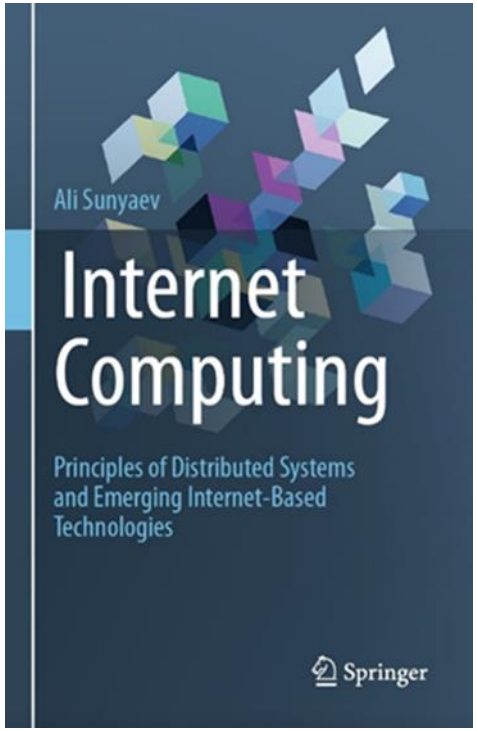
AI: Internet Computing

Lecture 4 — Internet Architectures

Learning Goals of the Lecture

- Understand how the exchange of messages over the Internet network infrastructure takes place
- Understand important Internet protocols, the functioning of domain names, and the routing of messages between senders and receivers of messages
- Learn about mechanisms that help to simultaneously deliver content to a large number of Internet users

Reference to the Teaching Material Provided



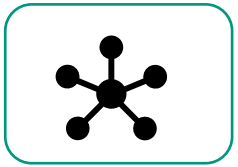
Chapter 4
Internet Architectures

Abstract
In order to explain how the Internet works, this chapter takes a closer look at the architecture that underlies the Internet, as well as at its architectural principles and mechanisms. After providing a brief overview of the Internet's history, this chapter examines today's core infrastructure and explains the role of Internet service providers. In addition, the essential mechanisms enabling Internet communication are explained, namely the Internet Protocol (IP) suite, IP addresses, the domain name system (DNS), as well as IP packet routing and forwarding. This chapter also explains how large content providers, like Google, Amazon, and Netflix, provide Internet users all over the world with efficient and reliable services by utilizing specialized content delivery networks. The description of four emerging architectural concepts that extend the established Internet architecture with more efficient and/or effective ways of providing innovative Internet services (i.e., software-defined networking, network function virtualization, overlay networks, and information-centric networking) conclude this chapter.

Learning Objectives of this Chapter
The main learning objective of this chapter is to understand the structure and the inner workings of today's Internet architecture. After studying this chapter, readers will understand the key organizations that keep the Internet running and how these

Recap: History of the Internet

History of the Internet

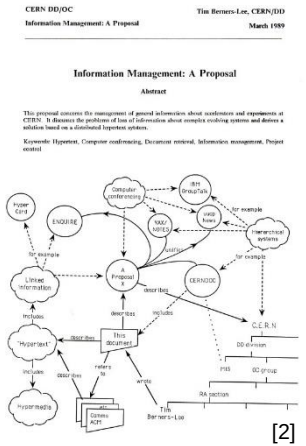


- In 1958 the Advanced Research Project Agency (ARPA) was founded by the US Department of Defense
  - To find solutions to protect the national long-distance communications networks from attacks of the Soviet Union
- In 1969, the first version of the ARPANET was finished
  - First permanent link between the University of California and the Stanford Research Institute
  - Later more links to other universities were added
- In 1971 the first email between two computers was send over the ARPANET
- 1974 design details for the Transmission Control Protocol (TCP) were published
- 1983 ARPANET adopts TCP and IP to its infrastructure

The World Wide Web

- Tim Berners-Lee invented the World Wide Web in 1989
- His basic idea was to merge the evolving technologies of computers, data networks, and hypertext into a powerful and easy to use global information system
- The document described a "hypertext project" called "WorldWideWeb" in which a "web" of "hypertext documents" could be viewed by “browsers”
  - Today known as “Hypertext Transfer Protocol (HTTP)
  - HTTP together with URI and HTML build the foundation of the World Wide Web


[1]


CERN DD/OC
Information Management: A Proposal

Tim Berners-Lee, CERN/DD
March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and discusses a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

[2]

Image source [1]: [Tim Berners-Lee] by CERN, July 10th 1994. Licensed under © CERN.
Image source [2]: [Proposal for the World Wide Web] by Cern, March 1989. Licensed under © CERN.

The World Wide Web

- By the end of 1990, the first Web server and browser was running at Cern
- With the first web page adress: http://info.cern.ch/hypertext/WWW/TheProject.html

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.

A replica of the NeXT machine used by Tim Berners-Lee in 1990 to develop and run the first WWW server, multimedia browser, and web editor.

Image source: [A replica of the NeXT machine used by Tim Berners-Lee in 1990] by Brice Maximilien, September 21st 2018. Licensed under © CERN.

The World Wide Web

Definition

The World Wide Web (WWW, or simply Web) is an information space (on the Internet) in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers.

T. Berners-Lee et al.

- It is important to emphasize that the terms Internet and WWW are not synonymous
- The WWW is just one service provided on the Internet
- HTTP is one of many protocols that enables data exchange over the WWW and can be used for communication on the Internet

Source: Berners-Lee T, Bray T, Connolly D, Cotton P, Fielding R, Jeckle M, Lilley C, Mendelsohn N, Orchard D, Walsh N, Williams S (2004) Architecture of the World Wide Web, Volume One. W3C

Uniform Resource Identifiers

- A Uniform Resource Identifier (URI) identifies a resource either by location, or a name, or both

- It has two specifications:
  - Uniform Resource Locator (URL)
    - Is a URI that identifies a resource and also provides the means of location of the resource by describing the way to access it
    - Example: http://www.ietf.org/rfc/rfc2396.txt
  - Uniform Resource Name (URN)
    - Is a URI that includes a name within a given space, but does not describe how to access the resource
    - Example: isbn:0-486-27557-4

Internet Network Infrastructure

Computer Networks

Definition

A computer network is a collection of computers and devices connected so that they can share information and services.

K.C. Mansfield, J.L. Antonakos

- Network devices either serve as hosts or facilitate relaying data between endpoints (e.g., modems, hubs, bridges, or switches)
- Two types of networks exist:
  - **Private Networks**
    - Requires users to obtain permission to gain access
    - Either manually by a network administrator or via password
  - **Public Networks**
    - Access is not restricted (e.g., the Internet)

Source: Mansfield KC, Antonakos JL (2009) Computer Networking for LANS to WANS: Hardware, Software and Security. Cengage Learning

Classification of Computer Networks

- Computer networks can be categorized by their purpose or geographical distance between nodes:
  - Local Area Network (LAN)
    - Connects computers and devices in a limited geographical area, like a building
    - Examples: Industrial plants, business offices, or domestic homes

Classification of Computer Networks

- Computer networks can be categorized by their purpose or geographical distance between nodes:
  - Local Area Network (LAN)
    - Connects computers and devices in a limited geographical area, like a building
    - Examples: Industrial plants, business offices, or domestic homes
  - Metropolitan Area Networks (MAN)
    - Interconnects LANs in a city or metropolitan area
    - Example: City networks

Classification of Computer Networks

- Computer networks can be categorized by their purpose or geographical distance between nodes:
  - Local Area Network (LAN)
    - Connects computers and devices in a limited geographical area, like a building
    - Examples: Industrial plants, business offices, or domestic homes
  - Metropolitan Area Networks (MAN)
    - Interconnects LANs in a city or metropolitan area
    - Example: City networks
  - Wide Area Networks (WAN)
    - Connects various locations such as campuses or offices
    - Example: the Internet

Overview Computer Networks

	Wide Area Network	Metropolitan Area Network	Local Area Network
Abbreviation	WAN	MAN	LAN
Purpose	Connects LANs with large geographical distances	Interconnects LANs in a city or metropolitan area	Connects computers and workstations within an office or home
Geographic Expanse	More than 50 km	5 to 50 km	Less than 5 km
Ownership	Private/Public	Private/Public	Private
Transmission Rates	Low	Medium	High
Propagation Delay	High	Medium	Low
Applications	Network between globally distributed enterprise branches, the Internet	City networks, several industrial facilities in close proximity	Industrial plants, business offices, university campuses, domestic homes

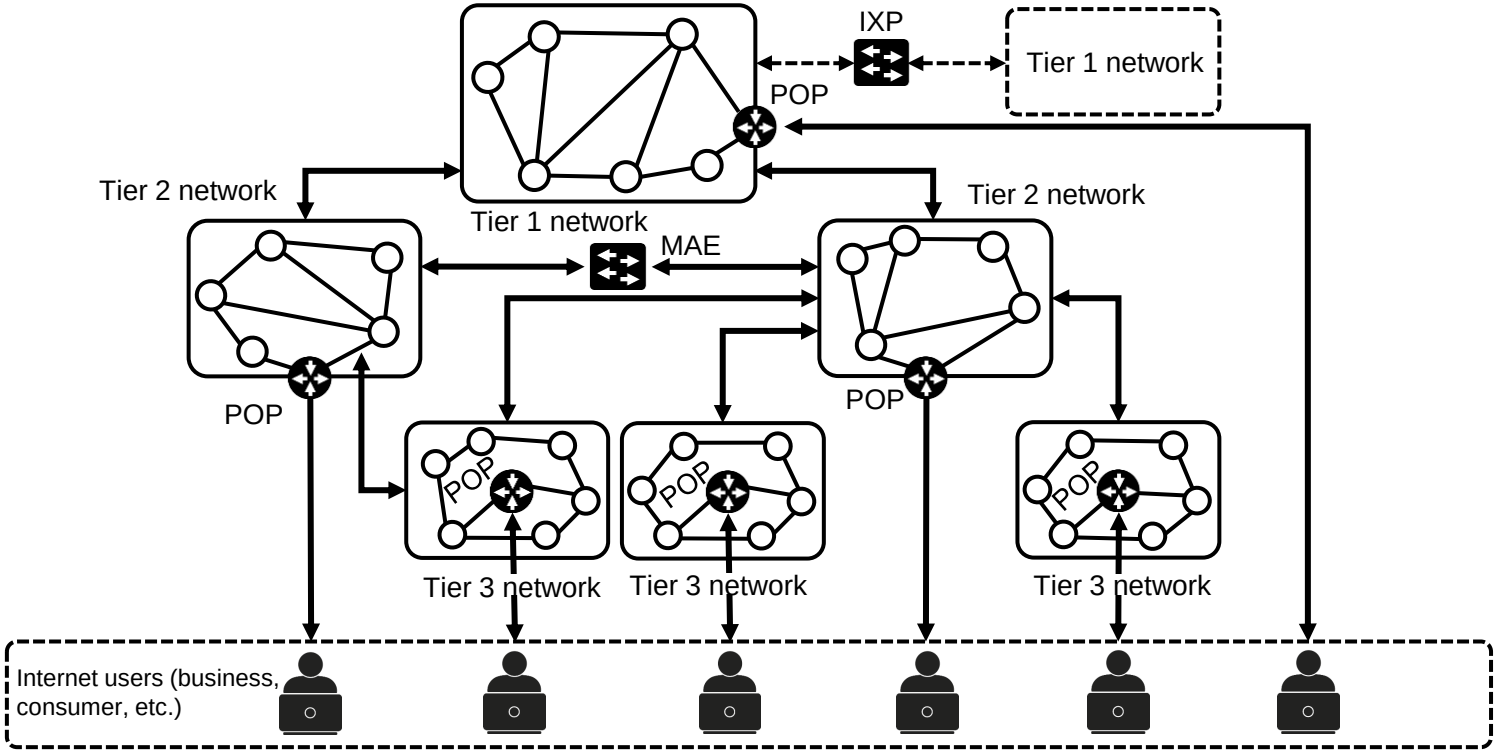
The Internet — A Network of Networks

Definition

The Internet can be defined as a public packet-switched wide area network that uses the TCP/IP protocol suite to interconnect computer systems across the world.

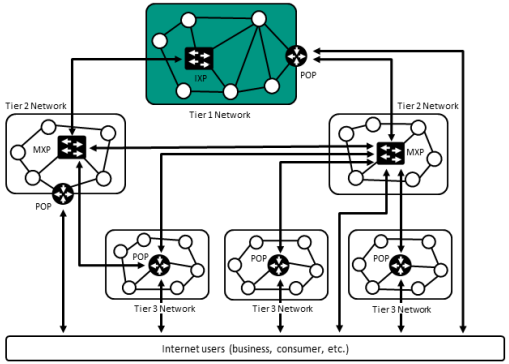
- The Internet provides a vast range of information resources and services:
  - Such as communication (e.g., electronic mail, telephony, and instant messaging),
  - or file transfer (e.g., file sharing, FTP, video, and audio streaming),
  - and the metaservice WWW.

The Internet — A Network of Networks



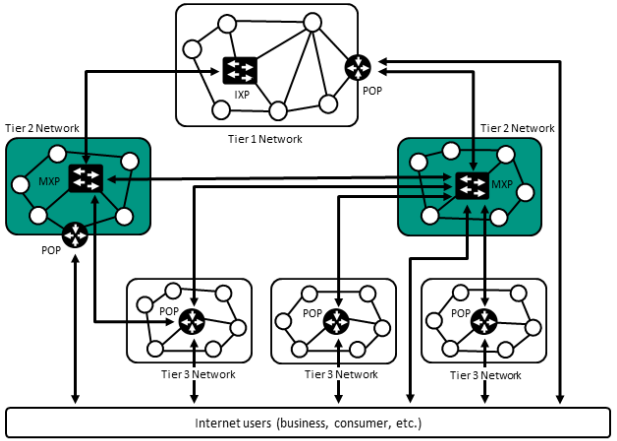
The Internet — A Network of Networks

- Tier 1 Networks:
  - Are operated by national telecommunication companies (e.g., Deutsche Telekom, AT&T); so-called Tier 1 ISPs (Internet service providers)
  - Are connected via high-speed optical fiber cable
  - Tier 1 networks exchange data directly with each other
    - This is called peering and does not involve any fees
    - Operational costs are covered by the providers involved in the exchange



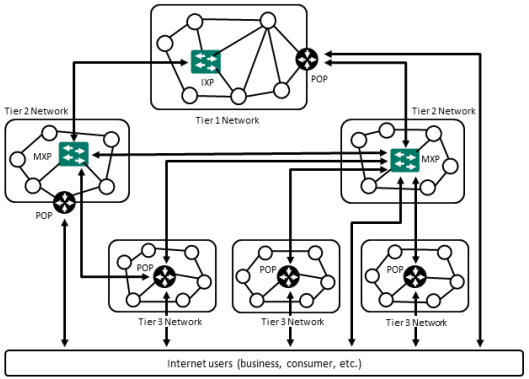
The Internet — A Network of Networks

- Tier 2 Networks:
  - They exchange Internet traffic through peering agreements and purchase Internet traffic from Tier 1 ISPs



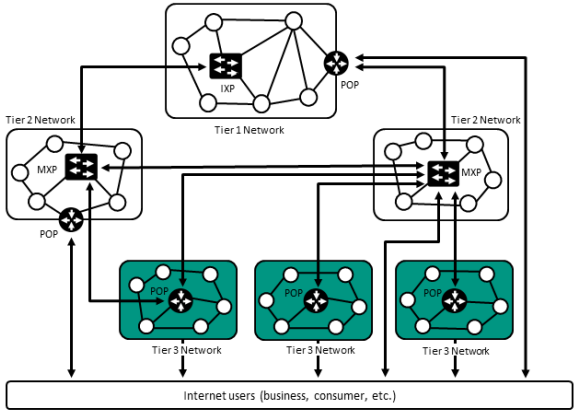
The Internet — A Network of Networks

- Tier 2 Networks:
  - They exchange Internet traffic through peering agreements and purchase Internet traffic from Tier 1 ISPs
  - Data exchange is done at neutral data centers, that is, data centers in shared use by multiple ISPs, called Internet Exchange Points (IXPs) or
  - Metropolitan Exchange Point (MXP)
    - MAEs are smaller versions of IXPs and typically link a set of tier 2 ISPs



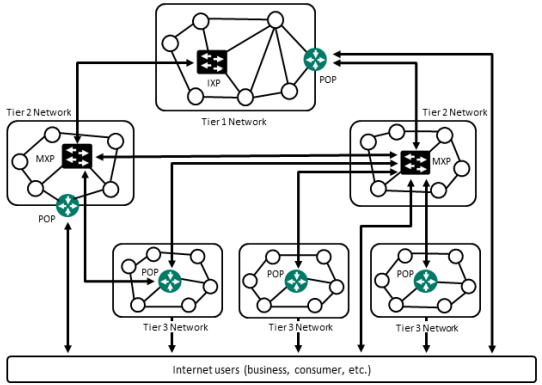
The Internet — A Network of Networks

- Tier 3 Networks (local ISP):
  - Delivers Internet access to residential homes and businesses
  - Strictly purchases Internet traffic from higher tier networks



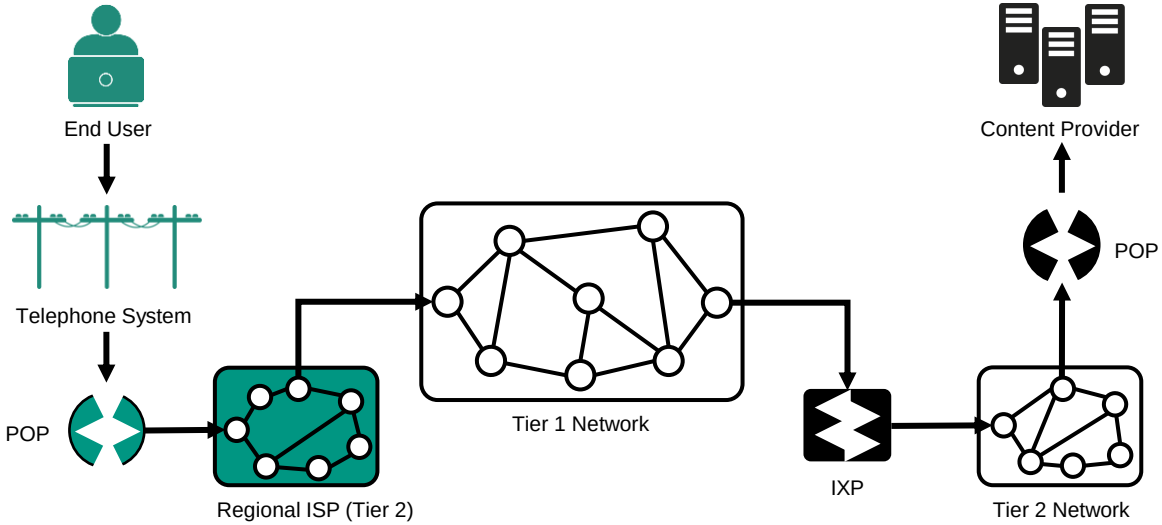
The Internet — A Network of Networks

- Tier 3 Networks (local ISP):
  - Delivers Internet access to residential homes and businesses
  - Strictly purchases Internet traffic from higher tier networks
  - Point of presence (POP):
    - Local access point of an ISP where the telecommunication lines from commercial or domestic buildings are connected to the ISP's network
    - Often located within the facility of a telecommunications provider responsible for the infrastructure to the customer



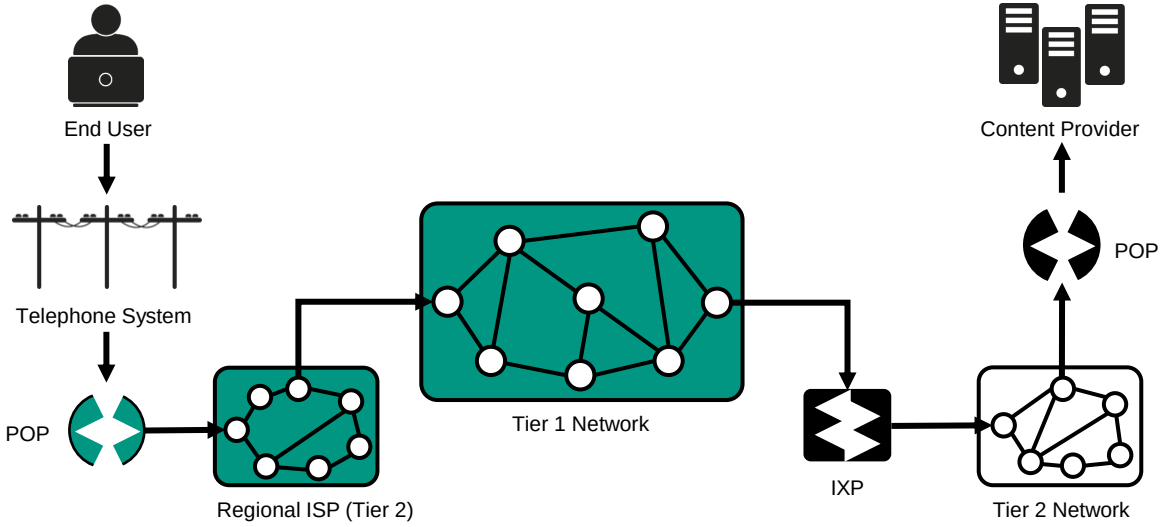
Internet Connection — Example

1) The transmission is initiated by the end user, whose computer is connected with the POP of his regional ISP



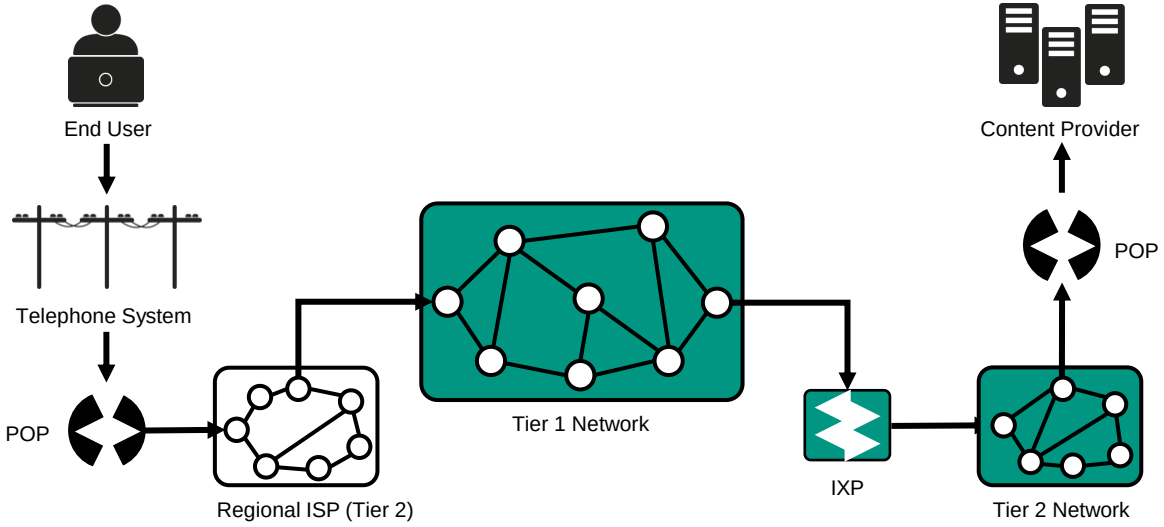
Internet Connection — Example

2) Destination is not in the regional ISPs network, therefore handover to a tier 1 ISP



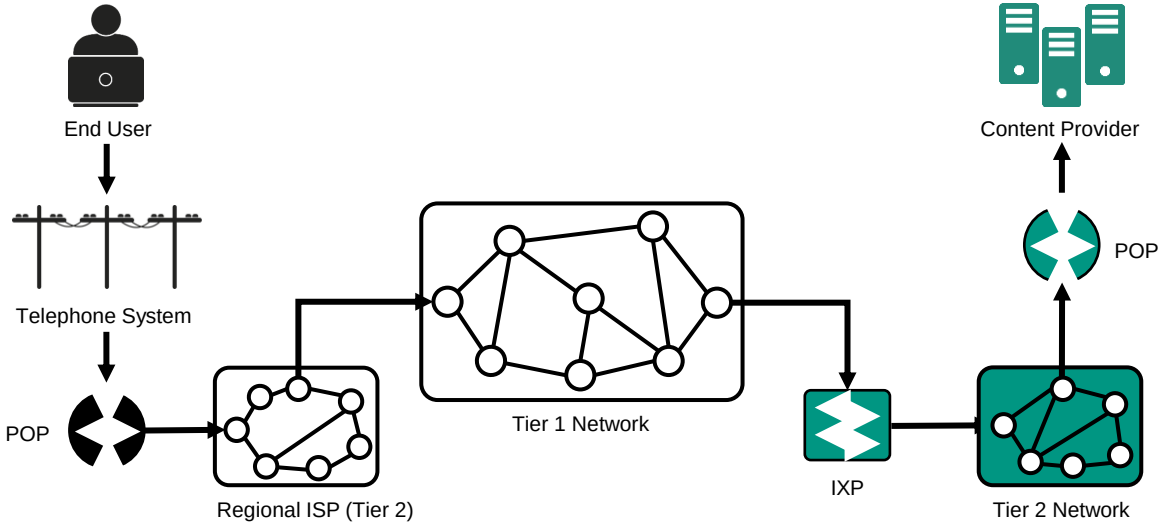
Internet Connection — Example

3) Tier 1 ISP establishs a connection with the destination tier 2 ISP



Internet Connection — Example

4) POP connects tier 2 ISP with the destination endpoint



Standardization of the Internet

- Several organizations standardize or regulate different aspects of the Internet or improve its stability and functionality
  - Internet Society (ISOC)
    - Provides organizational structure to support the process of Internet standard development
  - Internet Engineering Task Force (IETF)
    - Develops and maintains voluntary Internet standards (e.g., TCP/IP)
  - Internet Corporation for Assigned Names and Numbers (ICANN)
    - Responsible for the IP address space allocation and management of the domain name system
  - Word Wide Web Consortium (W3C)
    - Responsible for developing interoperable technologies for the WWW
    - Examples: HTML, XML, CSS, and SOAP

The Internet Protocol Suite

The Internet Protocol Suite

Definition

The **Internet protocol suite** is a set of protocols that enable communication over the Internet by specifying data transmission, addressing, and routing.

Baker

- What do we need an internet protocol suite for?
  - It governs how data is transferred from one system to another (through packetizing, addressing, and routing)
  - It is a collection of protocols that are designed to work together
  - It is the global standard for computer-to-computer communication

Source: Baker FJ (2009) Core Protocols in the Internet Protocol Suite. IEFT

The Internet Protocol Suite

Definition

The **Internet protocol suite** is a set of protocols that enable communication over the Internet by specifying data transmission, addressing, and routing.

Baker

- Most important protocols: TCP and IP
- TCP/IP provides end-to-end data communication over heterogeneous physical networks
- Key principles of TCP/IP:
  - Independency: no specific hardware and software requirements
  - Robustness: a built-in failure recovery mechanism provides reliable end-to-end communication

Source: Baker FJ (2009) Core Protocols in the Internet Protocol Suite. IEFT

Excursus: TCP/IP

- Transmission Control Protocol (TCP)
  - TCP is a reliable service which guarantees that all bytes are received in the right order
  - This is done by using positive acknowledgements (ACK) with re-transmission; receiver responds with an ACK for ever data packet he receives
  - Sender retransmits packets for missing ACKs after a given time
- Internet Protocol (IP)
  - IP is responsible for addressing host interfaces, encapsulating data into datagrams and routing data from a source host to a destination host

→ While IP handles actual delivery of the data, TCP keeps track of the data

TCP/IP Layers

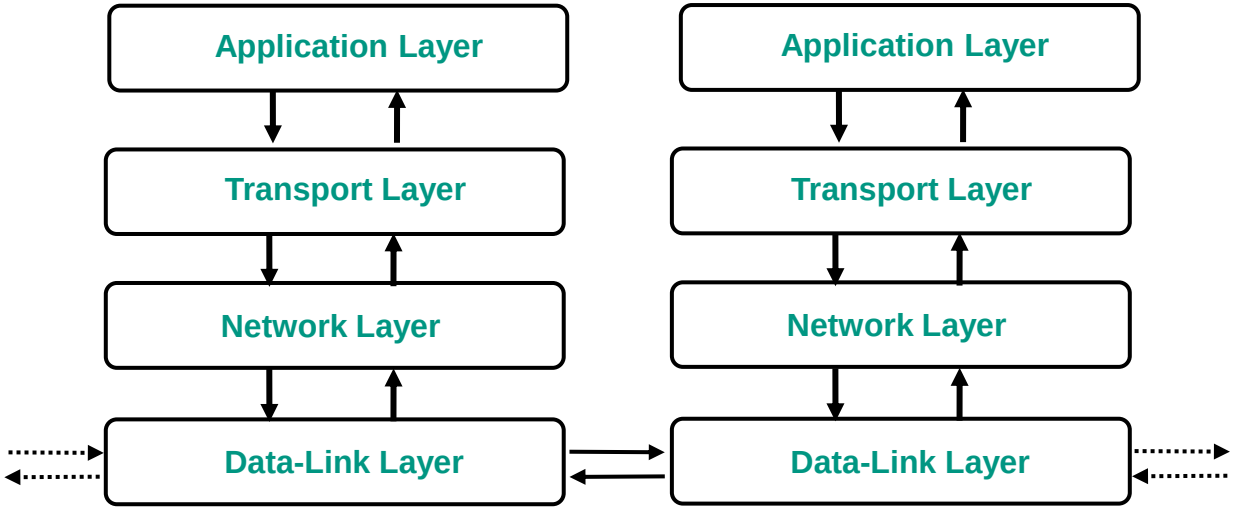
- It is possible to write a single protocol that takes data from one computer application and sends it to another one
  BUT: This is very inflexible
  → Therefore, **layered** protocol stacks are used!

- Each layer performs a specific function and communicates with the level above and below it

- All above the transport layer are services provided by the TCP/IP stack in the operating system
  → Therefore, developers are concerned only with interfaces in the application layer and the **transport layer**

TCP/IP Stack

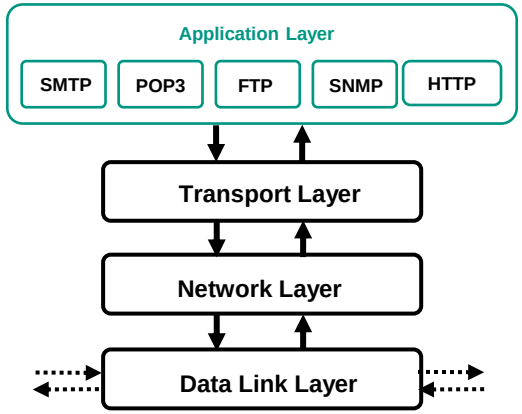
- TCP/IP comprises four abstraction layers:



TCP/IP Stack — Application Layer

The application layer...

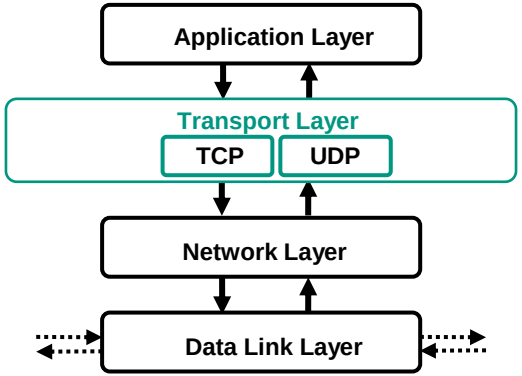
- … provides applications with standardized interfaces that allow to send data to other applications and receive data from them
- … makes use of lower layers and treats them as black boxes
- … contains all application-near protocols including **HTTP**, File Transfer Protocol (**FTP**), Post Office Protocol 3 (**POP3**), Simple Mail Transfer Protocol (**SMTP**), and Simple Network Management Protocol (**SNMP**)



TCP/IP Stack — Transport Layer

The transport layer...

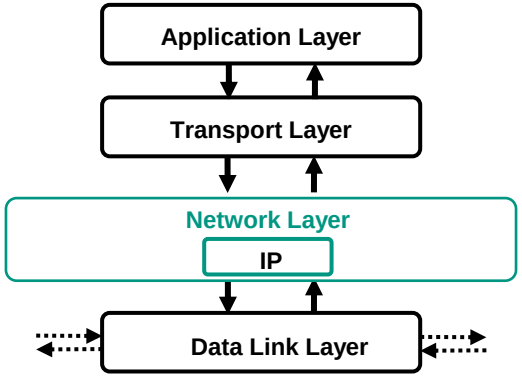
- ... provides end-to-end data transfer by delivering data from an application to a remote or a local host.
- ... **TCP** is well suited where data integrity is important
- ... **UDP** provides highly efficient but less reliable data transmission and has no error-recovery mechanism. It is therefore used for applications that need a fast transport mechanism and can tolerate the loss of some data (e.g., video streams)



TCP/IP Stack — Network Layer

The network layer (or Internet layer)...

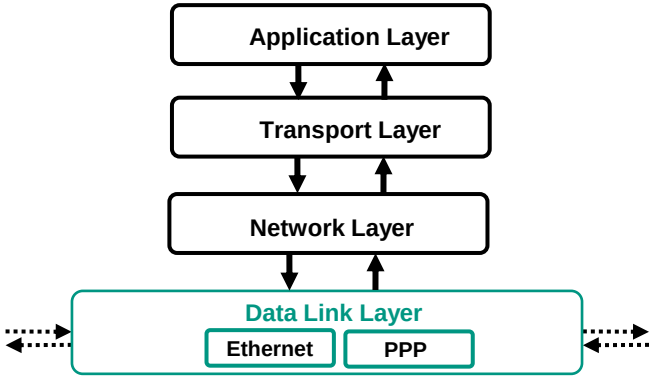
- … exchanges data in form of *datagrams* across network boundaries, provides a uniform networking interface and enables internetworking
- … defines the addressing and routing structures for the **TCP/IP** protocol suite. (**IP** is a connectionless protocol that provides a routing function that forwards data to a specific destination in the network that is identified by its unique IP address)



TCP/IP Stack — Data-link Layer

The data-link layer (network interface layer or physical layer)…

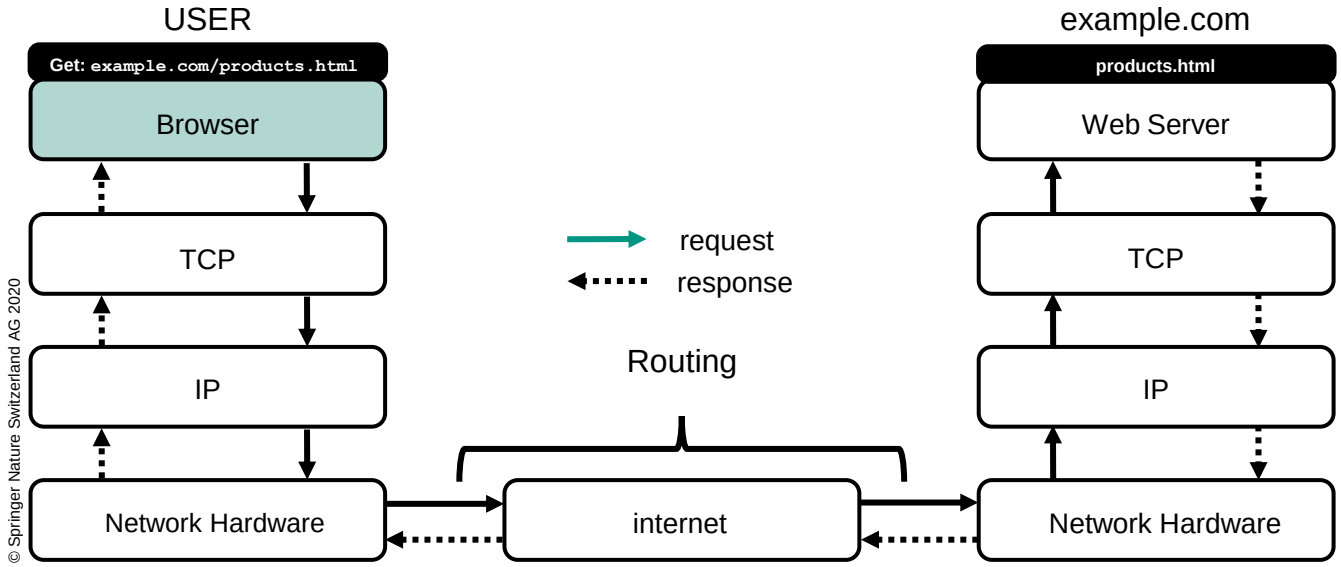
- … provides the interface to the actual network hardware
- … is the lowest layer because TCP/IP is designed to be hardware independent. As a result TCP/IP may be implemented on top of any networking technology
- … includes all protocols used to describe network topology and to move data between two different hosts (e.g., Ethernet)



The Internet Protocol Suite — Example

Example: Retrieving a webpage from a server

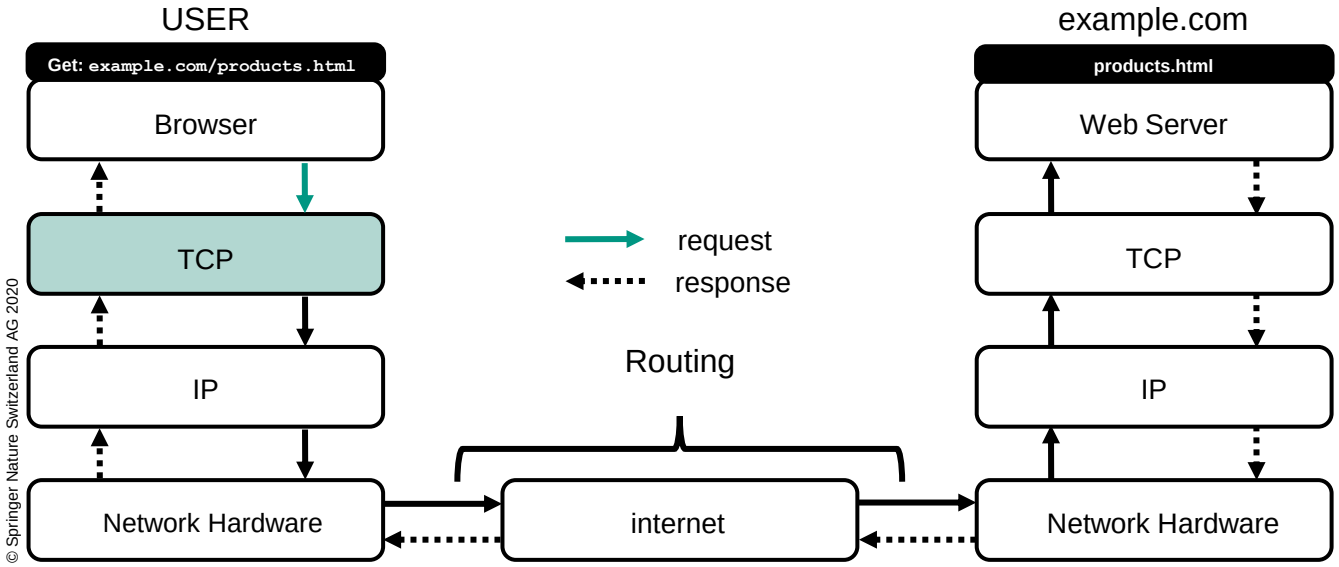
1) User opens the webpage with the URL *example.com/products.html*: User opens a Web browser application to communicate with Web server, which runs on application layer → application layer



The Internet Protocol Suite — Example

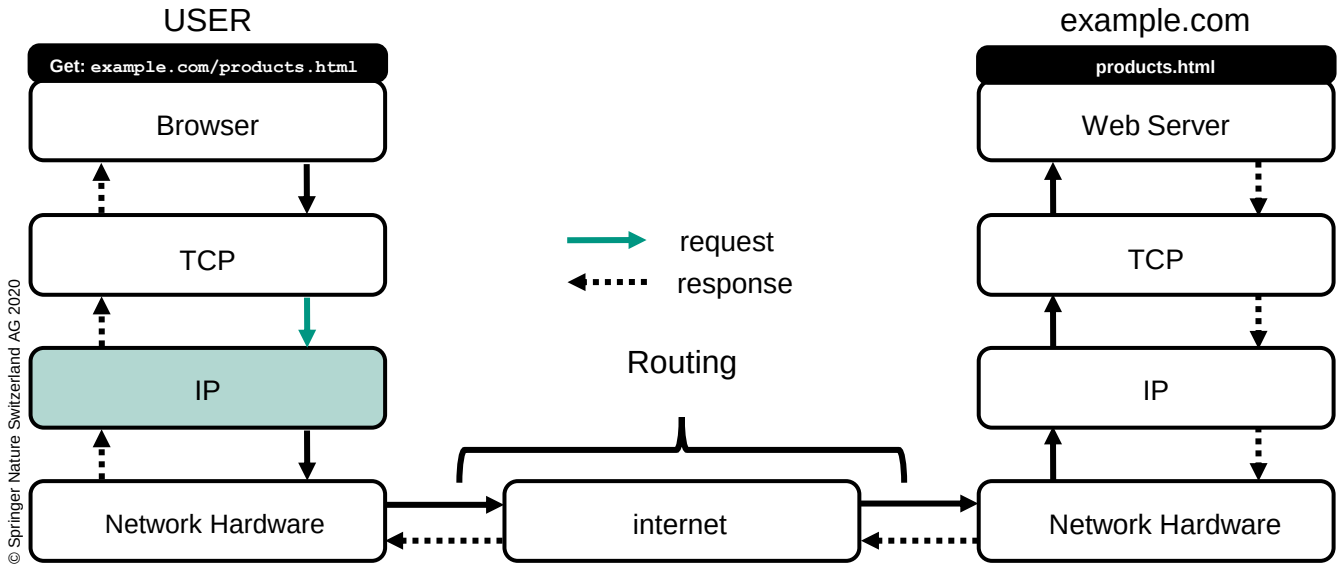
2) User hits *ENTER*: Browser program sends message *"Get: example.co.."* to the TCP program → transport layer

3) TCP program adds information about the address of the source and the host and adds an error checksum



The Internet Protocol Suite — Example

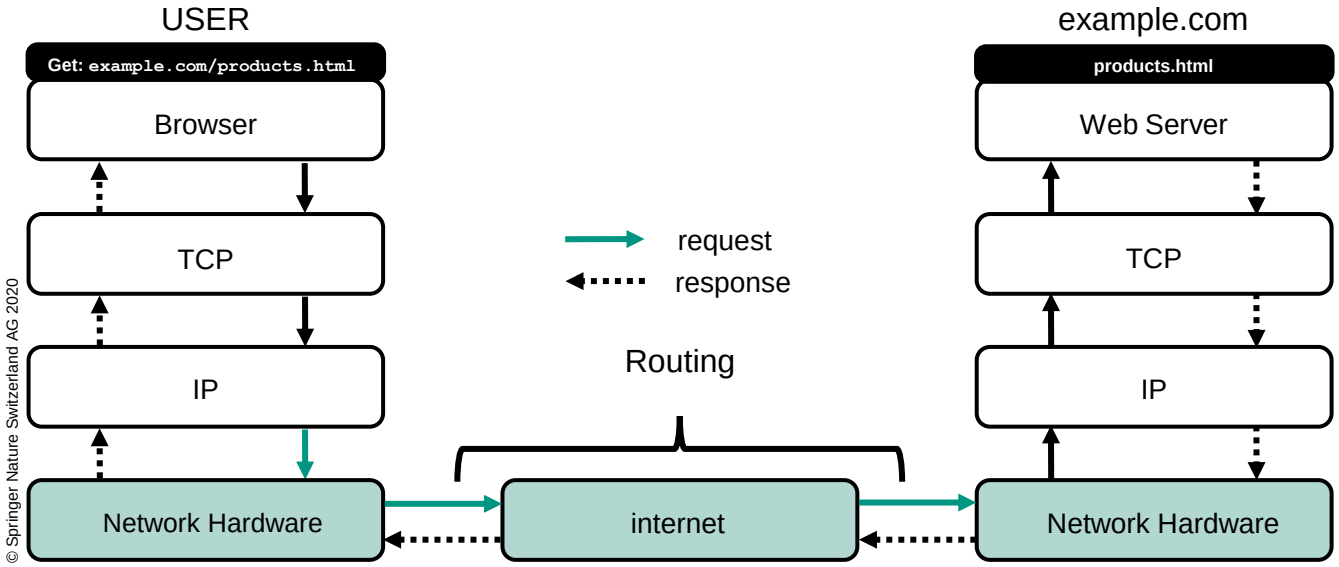
4) TCP program passes information to the IP program → network layer
5) IP program adds information about the source and destination IP addresses



The Internet Protocol Suite — Example

6) The message gets transmitted through the Internet by hardware, software and firmware at the bottom of the protocol stack → data-link layer

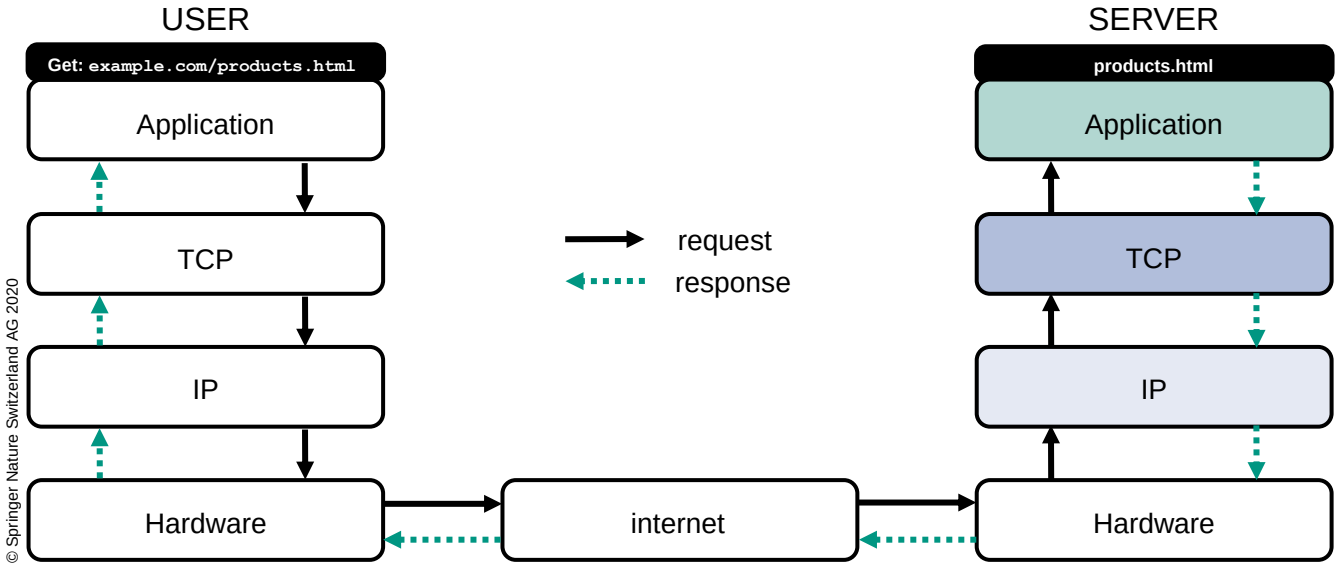
7) The message traversers the servers same protocol stack in reverse order



The Internet Protocol Suite — Example

8) Once the message reaches the application layer, the server will prepare a response message that includes the document *“products.html”*

9) And sends it all the way back



IP Addresses

Definition

An **IP address** is a unique string of numbers separated by full stops that identifies each computer using the Internet Protocol to communicate over a network.

Stevensen & Waite

- What do we need IP addresses for?
  - Having an IP address allows a device to communicate with other devices over an IP network such as the Internet
  - IP addresses provide an identity to a network device which is essential for transferring data: It is similar to sending a packet to a business address: You have to attach a *specific* address (IP address) to the packet or look it up in a phonebook (DNS Servers) to make sure it arrives at the *exact* destination
  - IP addresses are assigned by central authorities to ensure uniqueness

Source: Stevenson A, Waite M (2011) Concise Oxford English Dictionary: Book & CD-ROM Set. OUP Oxford

IP Addresses

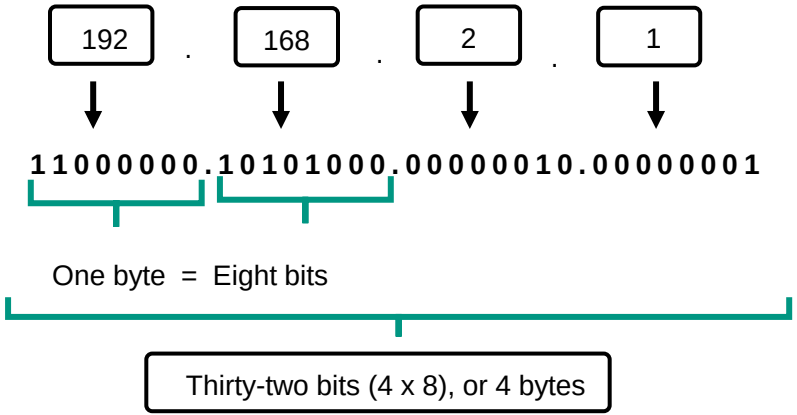
Two versions are in use simultaneously: **IPv4** and **IPv6**

- IPv4 (Internet Protocol version 4) (1983)
  - 32-bit long binary string
  - Specifies 2^{32} (4,294,967,296) unique addresses
  - Pool of addresses became to small due to rapid growth of the Internet

- IPv6 (Internet Protocol version 6) (1995)
  - 128-bit long binary string
  - Specifies 2^{128} (approx. 3.403 x 10^{38}) unique addresses

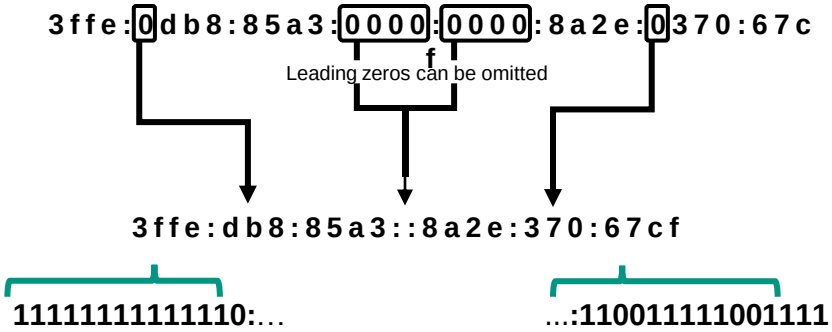
IPv4

- Four-part decimal number, separated through decimal points
- Example: 192.168.2.1 (each of the numbers is an eight-digit binary number called octet)
- Value of individual bytes: 0 - 255 (00000000 – 11111111)



IPv6

- Eight hexadecimal numbers, separated through colons
- Each group has four hexadecimal digits that represent 16 binary digits (referred to as a hextet)
- Example: 3ffe:0db8:85a3:0000:0000:8a2e:0370:67cf
- Hexadecimal digits are not case-insensitive, but IETF recommends to use lower-case letters



Packet Switching and Routing

We now know **where** we send our data (to IP addresses).
But: **How** does the data find its way from A to B over the Internet?

→ Through Packet Switching and Routing

Packet Switching

Definition

Packet Switching describes a switching and transmission technology which splits complete messages into smaller packets. These packets can be transmitted along different lines of a network and they are re-assembled into the original message by the receiving host.

Jordana

- TCP/IP uses Packet Switching to transmit data
- **Why** do we use Packet Switching instead of sending the entire message in one large packet?
  - It optimizes the use of the available **channel capacity** in networks, such as computer networks.
  - It minimizes the **transmission latency** (the time it takes for data to cross the network).
  - It increases the **robustness** of the network

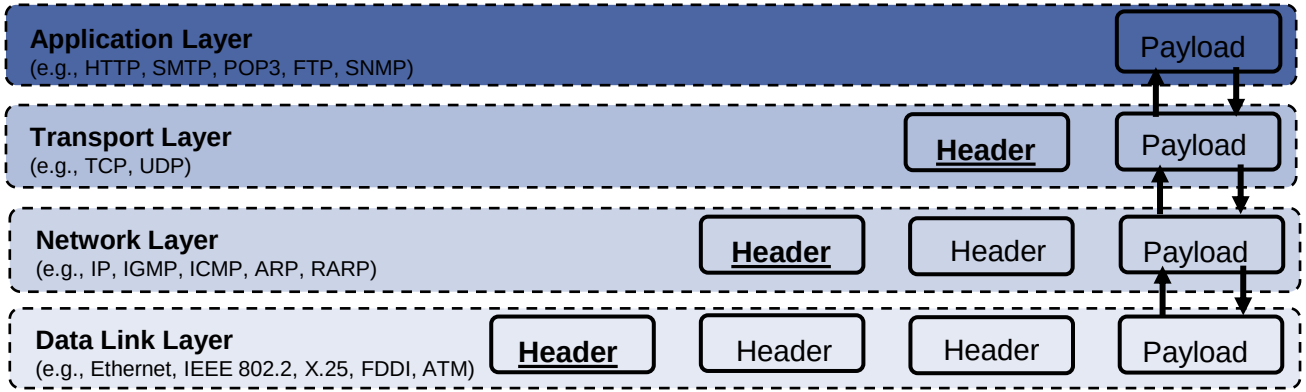
Source: Jordana J (2002) Governing Telecommunications and the New Information Society in Europe. Edward Elgar Publishing, Incorporated

Packet Switching

- The best-known use of Packet Switching is the Internet and local area networks
- The structure of packets depends on the applied protocol (Ethernet is the most common)
- IP packets consist of a header and a payload, so called Datagram
  - Header: keeps information about the packet, the service, and other transmission-related data (e.g., IP address of sender and destination, sequence number of the packets)
  - Payload: holds the actual carried data

Packet Switching

- An IP packet is usually enveloped by multiple layers, one for each protocol used for a particular connection and each with its own header (like an onion)
- Each layer treats information from above layers as data
  → the process of preserving the data while attaching a new header is known as encapsulation



Router

Now that the data is broken down into numerous smaller parts that were packaged in specially formatted units called packets, these packets need to be **transferred**

→ Packets are being transferred over **routers** through **routing!**

Router

Definition

On a network, a device that determines the best path for forwarding a data packet toward its destination. **The router** is connected to at least two networks and is located at the gateway where one network meets another.

Grance et al.

- Routers are able to **forward packets** beyond the borders of a network (important for the multi-network architecture of the Internet)
- Routers check whether incoming packets contain errors
  - If they contain an error that cannot be fixed the router discards the packet and sends an error message back to the sender
  - If there is no error, the destination address of the packet is read from its header and the packet's next routing target is determined based on routing algorithms (e.g., Distance Vector, Link State and Path Vector routing)

Source: Grance T, Hash J, Peck S, Smith J, Korow-Diks K (2008). Security Guide for Interconnecting Information Technology Systems. NIST

Domain Name System

Definition

The **Domain Name System (DNS)** is a hierarchically structured, distributed set of databases that map IP addresses to corresponding domain names.

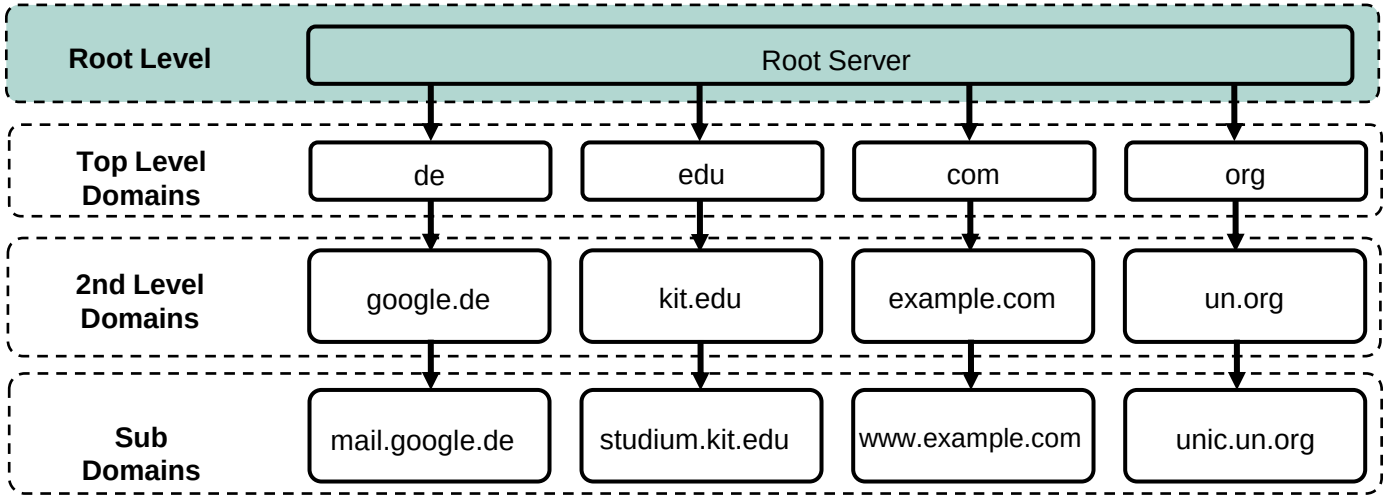
Chandramouli

- IP addresses are not practical → IETF proposed Domain Name System (DNS)
- DNS allows Internet users to visit a website by typing the domain name rather than the IP address
- Example: Use https://kit.edu instead of https://[2a00:1398:b::8d03:8006] to access the KIT website

Source: Chandramouli R, Rose S (2006). Challenges in securing the domain name system. IEEE Computer Society 4 (1):84-87

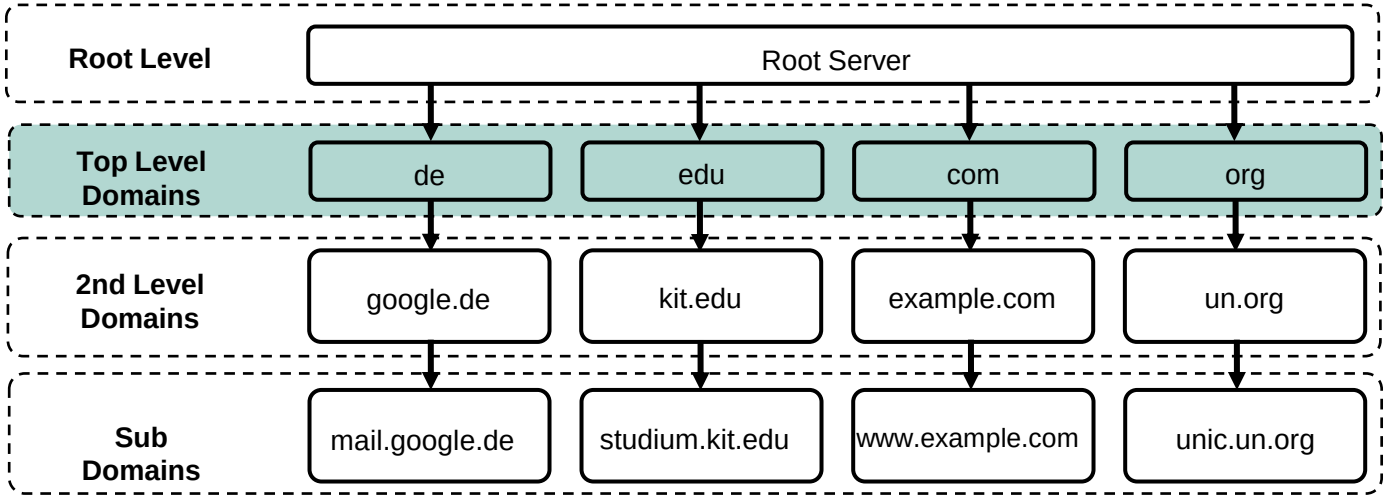
Domain Name

- A domain name consists of one or more labels, separated through decimal points
  - Each label specifies a subdomain of the domain
  - Subdomains are organized hierarchically in a tree-like structure, starting from the nameless DNS root domain



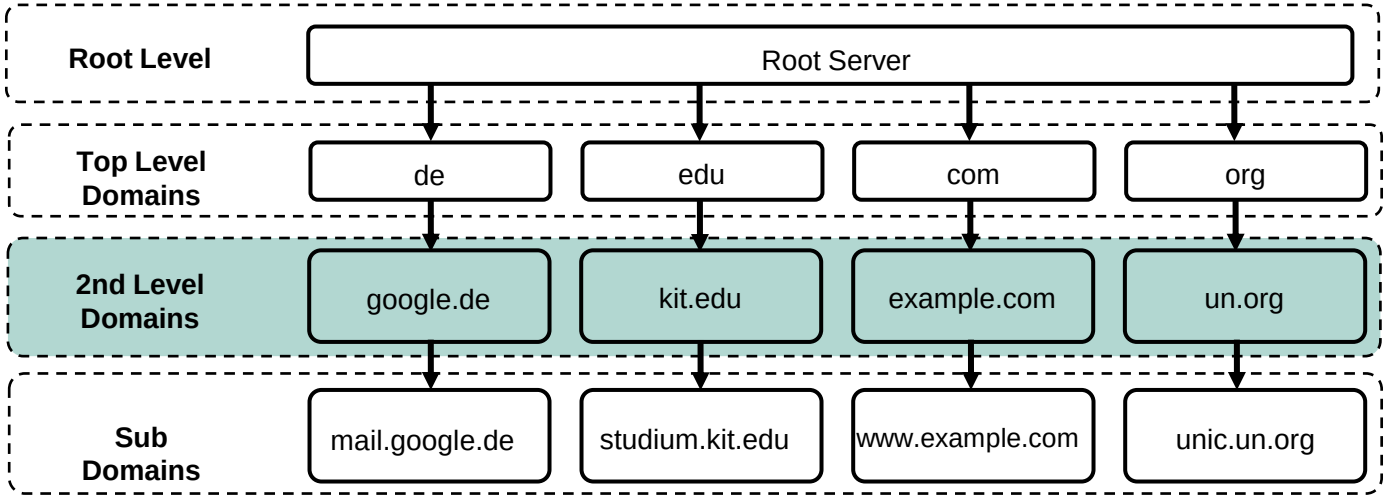
Domain Name

- Top-level domains
  - Generic top-level domains: com, info, net and org
  - Country code top-level domains: de, fr, ca, …
  - Sponsored top-level domains: edu, gov, jobs, ...



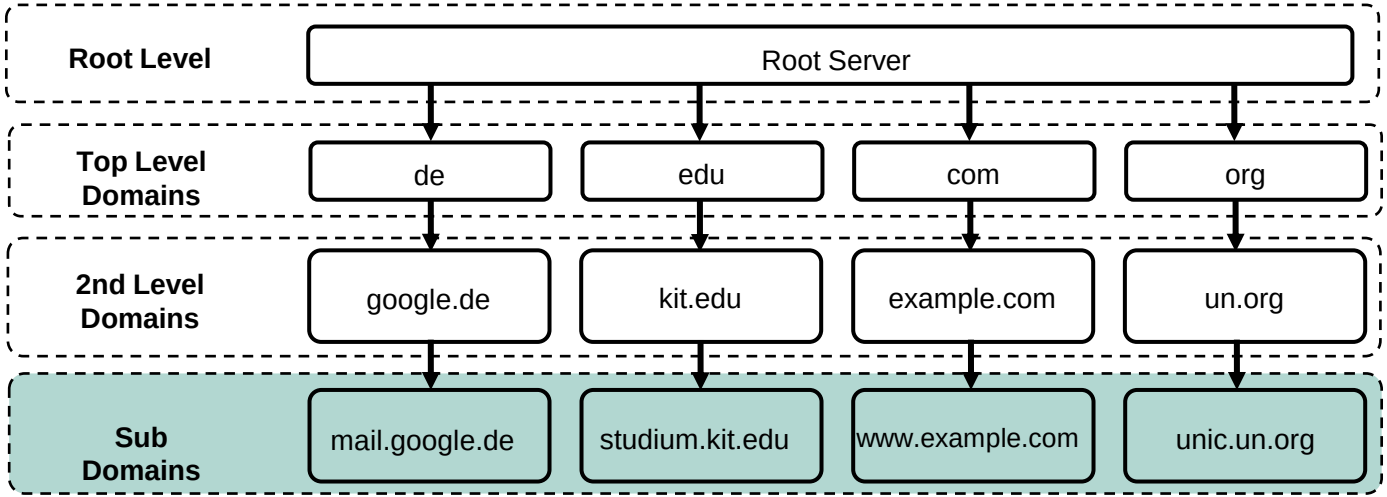
Domain Name

- 2nd level domain
  - Open for reservation by organizations and end-users (e.g., KIT)



Domain Name

- Sub domain / 3rd level domain
  - Not mandatory
  - Used to specify a certain server inside an organization
    - For example, www, conventionally links to a Web server providing WWW services



Domain Name Systems

- A distributed network of **DNS servers** manage, maintain, and process domain names and their associated records
- All information is stored in **DNS translation tables** on numerous **DNS servers** around the world
  - DNS servers only store records associated with domain names of a particular zone
  - DNS servers are updated regularly (between a few hours and a few days)

Domain Name Systems

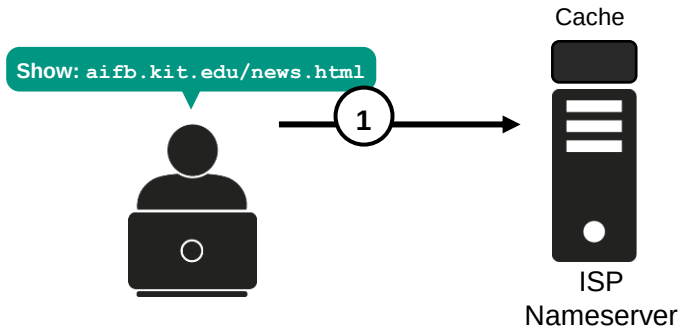
We know the communication between two endpoints of the Internet requires each endpoint to have a unique IP address.

How does an endpoint know which IP-address is assigned to a domain name?

→ DNS lookup

DNS Lookup — Example

1) User wants to visit a website using its domain name, so the browser sends a request to a known DNS server

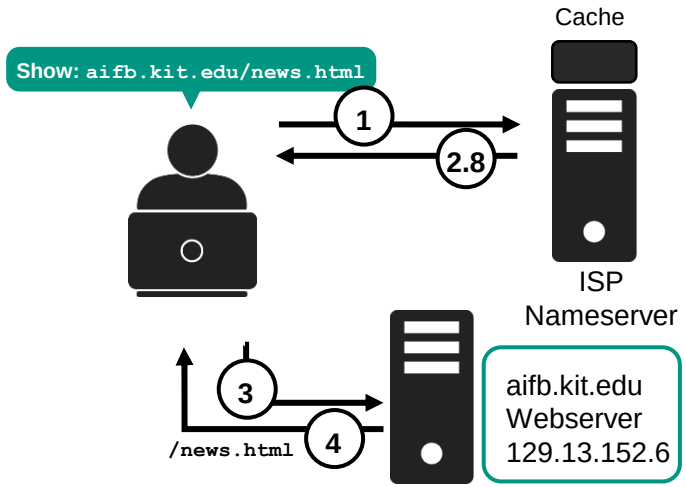


DNS Lookup — Example

2.8) In case the domain name is known to the server, as it is stored in the cache, it will directly provide the relevant DNS record

3) The right server is contacted

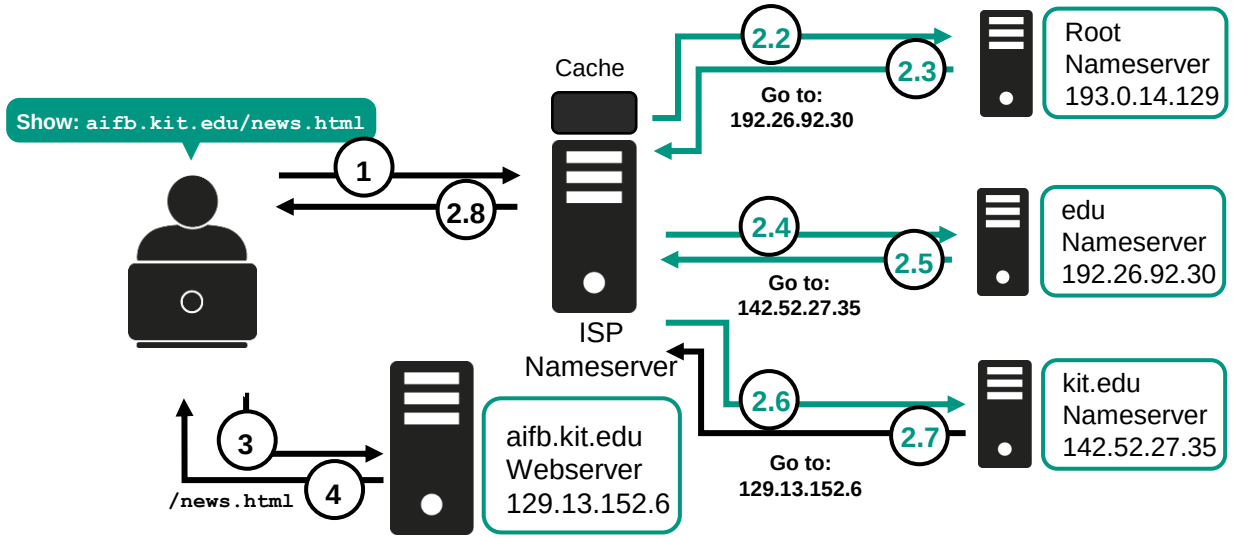
4) User receives the requested *"/news.html"*



DNS Lookup — Example

2.2) If a DNS server does not know the requested domain name, it will refer to the DNS server for the root-level domain

2.3 – 6) The root-level DNS server refers to top-level DNS servers, which in turn refer to 2nd-level DNS servers until the IP is known

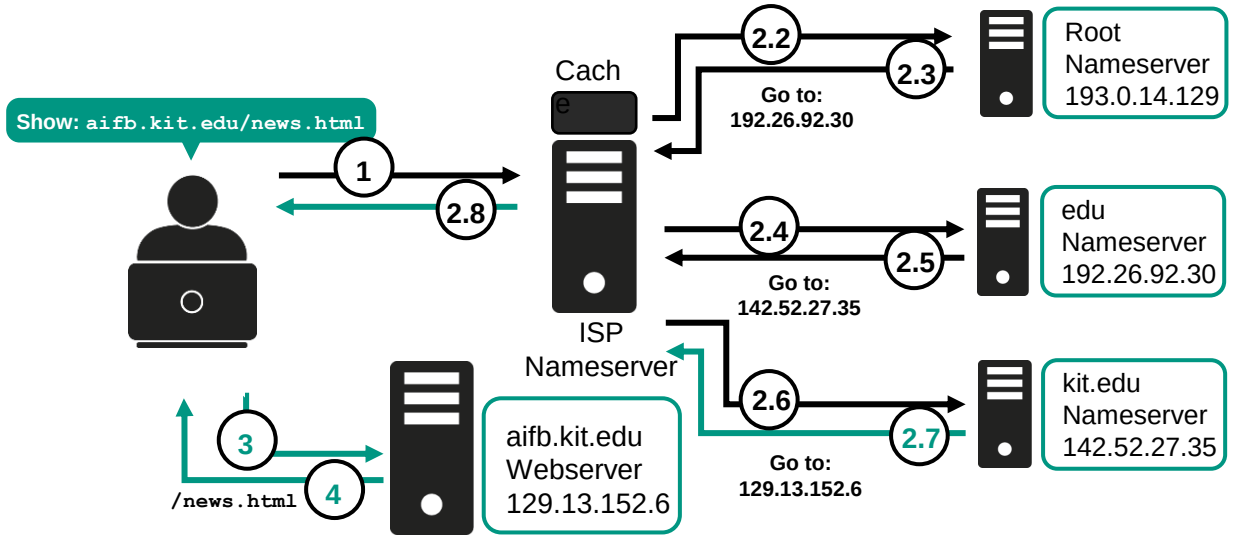


DNS Lookup — Example

2.7) The DNS server on the 2nd Level provides the IP address of the requested domain name

3) The right server is contacted

4) User receives the requested *"/news.html"*



Content Delivery Networks

Content Delivery Networks (CDNs)

- **Reliable** and **fast delivery of content** over the Internet is a **challenging task**
  - Unpredictable demand peaks that exceed available server resources
  - Large geographical distances between the content providers and consumers

- **Load balancing spreads** the overall hosting burden of one server or content provider across **multiple devices** (no single point of failure)
  - Intuitive approach: single data center that hosts multiple redundant servers
  - But: typically **same location** → **single point of failure** → do little to help with problems caused by **network issues**, **large distances,** or **outages**

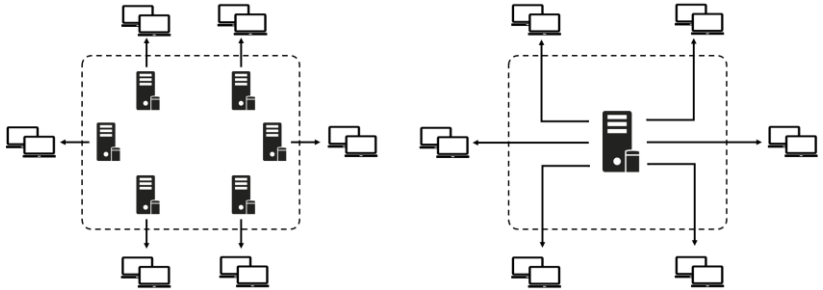
Content Delivery Networks (CDN)

Definition

Content Delivery Networks (CDNs) are a collection of network devices that are controlled by a common management infrastructure with the main purpose of delivering content (e.g., websites, videos) more effectively to clients over the Internet.

IETF

- Have been deployed in increasing numbers in recent years
- It is expected that by 2022 up to **72%** of the global Internet traffic will be handled by **CDNs**



CDN Scheme of Distribution

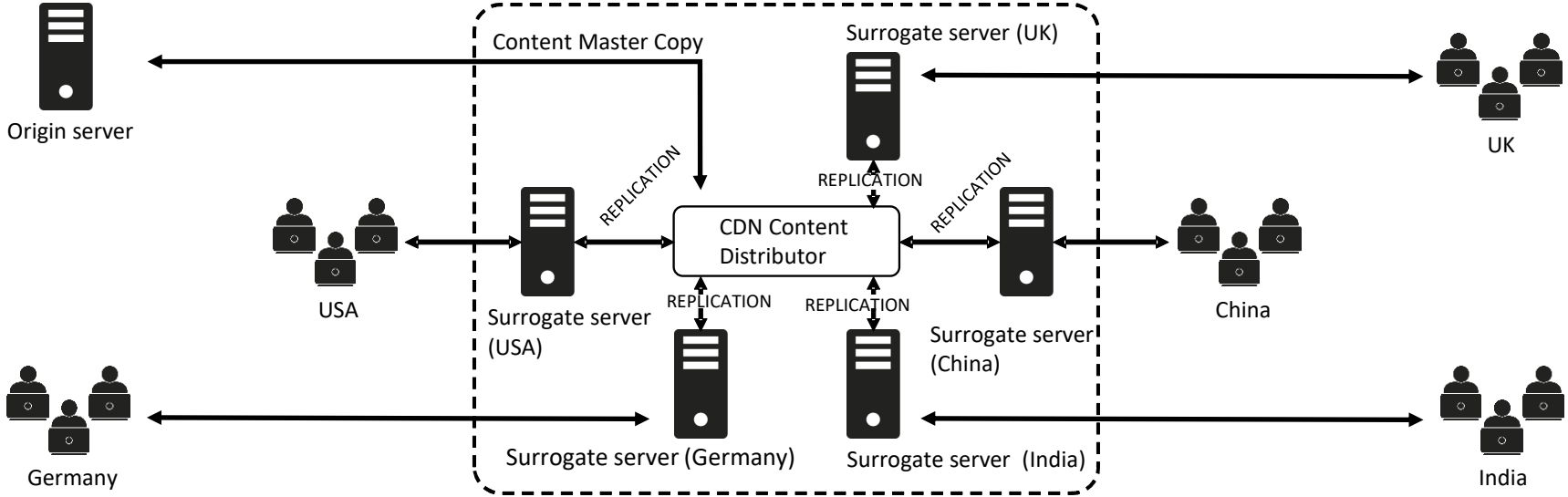
Single Server Distribution

Source: IETF (2003) A Model for Content Internetworking (CDI).
Cisco (2018) Cisco Visual Networking Index: Forecast and Trends, 2017–2022.

Content Delivery Networks (CDN)

- Basic principle is to move content to network locations **closer to content consumers**
- **Reduction of distance** decreases latencies, the risk of connection interruptions, and improves transmission speed
- This is done by **replicating** the content onto multiple content delivery servers in **different geographical locations** (**surrogate servers**)
- Clients' content requests are **automatically routed** to the surrogate servers (process often invisible for the requesting clients)
- CDNs are often operated by specialized **network service providers** such as Amazon Web Services (AWS)

Content Delivery Networks (CDN)



Content Delivery Networks (CDN)

- When a client requests a single content item from a CDN the request is directed to the best suited surrogate server holding a copy of the item
- For example, a CDN could operate one surrogate server per continent
- Surrogate servers and their management infrastructure form the **content delivery infrastructure**
- The **request-routing infrastructure**
  - Handles the steering and directing of content requests from a client to the right surrogate servers
  - Selection of the **“best-suited” surrogate** usually means that the item will be served by the server that promises the shortest delivery time with appropriate integrity and consistency
  - This selection usually requires **dynamic information** about **network conditions** and **load on the servers**

Content Delivery Networks (CDN)

- The **distribution infrastructure**
  - Handles activities concerning the **distribution of content** within the CDN.
  - All content is first published on an origin server: The “master copy“ of a content item is then replicated onto one or more surrogate servers
  - This distribution can happen either in anticipation of a surrogate server receiving a user request (**pre-positioning**) or in response to a surrogate server receiving a response (**fetching on demand**)
- The **accounting infrastructure**
  - Responsible for measuring and recording of the networks content distribution and delivery activities
  - These records are used to calculate the service fees for the client (a content provider)

Types of CDN Infrastructures

- Content provider and content consumer might not be the same entity
- **Private CDN**
  - A globally operating cooperation can build their own CDN to deliver content to their different subsidiaries across the globe
- **Federated CDN**
  - Based on infrastructure that is operated by multiple content or service providers
  - Participating providers pool their existing resources into a single delivery network
  - Two possible deployment approaches:
    - Bilateral approach: every participating provider directly interconnects with every other provider in the network
    - Exchange approach: every participating provider connects to a central hub that provides internetworking functionalities (e.g., routing)
  - Enables smaller content or service providers to **compete with** larger CDNs

Types of CDN Infrastructure

- **Peer-to-peer (P2P) CDN**
  - Surrogate servers are either partly or completely substituted by the network‘s clients (i.e**.,** **peers**), which both provide and consume the content
  - Creates a **mesh network** consisting of users who want to access the same content
  - Coordinates its clients so that they send chunks of the item to each other

- P2P CDN can actually **perform better** with more clients accessing the content because the network load is distributed among more clients
- While this approach limits the content provider‘s **control over the delivery process**, it heavily reduces the setup and operational costs of the CDN

CDN Summary

→ Each CDN type has its own **set of benefits** and **suitable scenarios** and the optimal approach that best suits the purpose needs to be **selected by the content provider**

→ However, all approaches have in common that they offer **smooth, reliable, and fast service delivery** for content that is popular, exhibits fluctuating request patterns, or is demanded by a geographically distributed audiences

Software Defined Networking

Software Defined Networking (SDN)

- SDN is an approach to networks that **separates** the **network's control and forwarding layers**
- This separation enables the network control to be **directly programmed** and the **underlying infrastructure to be abstracted** for applications and network services
- Problem with traditional network architectures:
  - **Reliance** of network functions on purpose-built hardware
  - Decentralized and complex infrastructures with **large administrative overhead**

  → Changes in traditional network architectures involve time-consuming process adjustments and **costly investments**
- Goal of SDN is to **overcome** this challenge by enabling application and network engineers and administrators
  - to **respond quickly** to changing business requirements via centralized controls
  - to improve **network performance and monitoring**

Software Defined Networking (SDN)

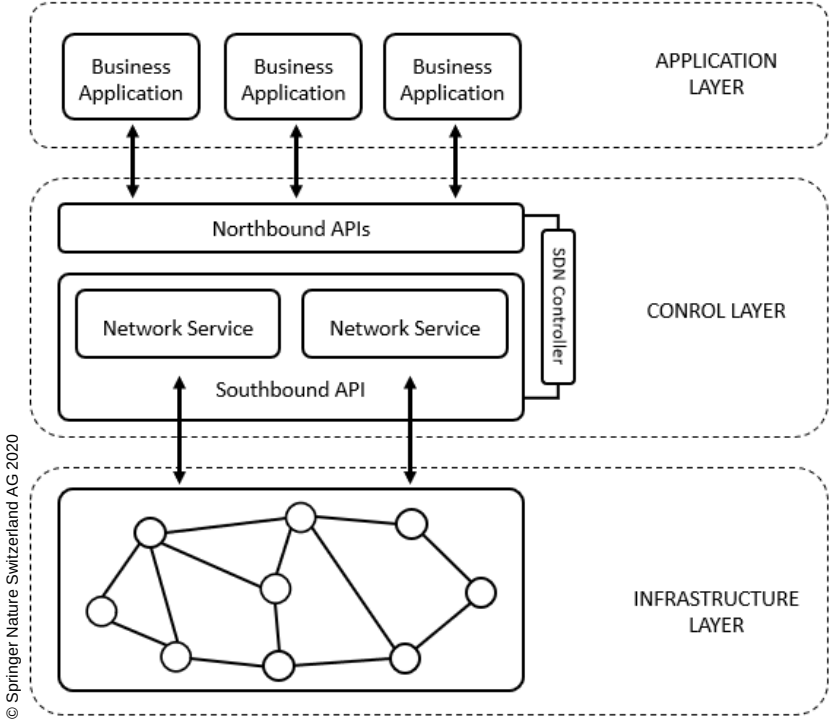


Image source: Adapted from Open Networking Foundation (2012). Software-defined networking: the new norm for networks. https://opennetworking.org/sdn-resources/whitepapers/software-defined-networking-the-new-norm-for-networks/. Accessed 27 April 2022

Software Defined Networking (SDN)

- SDN **centralizes the control** of the network in one network component
  - By separating the forwarding process of network packets (data plane) from the routing process (control plane)
  - Thereby moving the control logic to off-device computer resources
- Numerous different SDN implementations have in common that they all contain an **SDN controller**, **southbound APIs**, and **northbound APIs**
- The **SDN Controller** acts as the brain
  - Incorporates the whole intelligence
  - Offers a centralized view of the overall network and enables network administrators to dictate to the underlying systems how the forwarding plane should handle network traffic
  - Communication between the different layers is enabled by **application programming interfaces (APIs)**

Software Defined Networking (SDN)

- The **southbound APIs** relay information from the controller “down” to the infrastructure components
  - Control the path of network packets across network switches
  - First real standard was the **OpenFlow protocol** (first introduced in 2011)
  - Remains one of the most common protocols
  - Other approaches have emerged as well (Open Network Environment by Cisco Systems, Nicira's network virtualization platform)
- The **northbound APIs** enable communication between the SDN controller and the applications and business logic “above”
  - Applications can use the APIs to **program the network** according to their needs
  - Applications **use the (customized) network** through services provided over the northbound APIs
  - Currently no particular standardized protocol for northbound APIs exists
  - Often implemented by using **RESTful service interfaces**

Software Defined Networking (SDN)

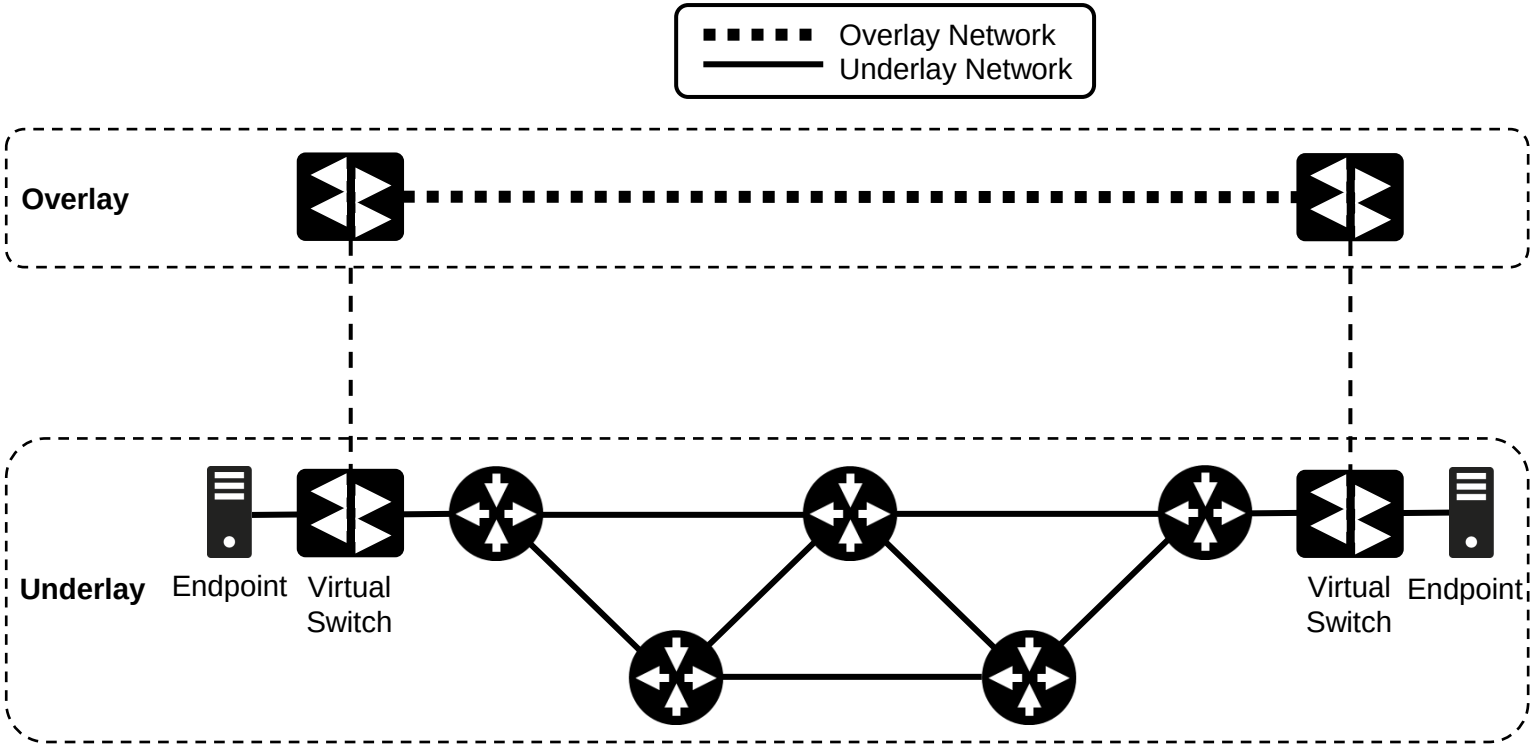
- A major advantage of the API-based approach in SDN is that the resulting **loose coupling** between the three layers enables a **complete virtualization and dynamic allocation** of network and service functions onto an arbitrary physical infrastructure
- SDN promises numerous benefits including:
  - On-demand provisioning of network traffic
  - Automated load balancing
  - Streamlined physical infrastructure
  - Ability to monitor and precisely adapt network resources to exactly match application and data needs
- One practical example for where these benefits can shine is **network security** (network anomalies can be detected much faster with the centralized control layer)
- Nonetheless, the centralization of network controls (i.e., creation of a single point of failure) also comes with downsides in terms of reliability, security, and scalability

OVERLAY NETWORKS

Overlay Networks

- Overlay Networks (also called SDN overlays) are **virtual networks** of **nodes** and **logical links**
- Use software to create layers of network abstraction that can be used to run multiple separate, discrete virtualized network layers on top of a physical network
- Main purpose is to **enable new services** or **functions** without having to reconfigure the entire network design
- Most of today's overlay networks run **on top** of the **public Internet**
- For example, virtual private networks (VPN), peer-to-peer networks, and VoIP services (e.g., Skype) use this technology
- The **WWW itself** can be seen as an **overlay network**

Overlay Networks



Overlay Networks

- Overlay Networks are created by defining **two or more endpoints** and creating a **virtual connection** between them
  - Endpoints can be physical locations (network ports) or logical locations.
  - Somewhat similar to the **phone system** with endpoints designated by an identification tag or number
  - Connection is created using an **overlay node software architecture** that is shared between all overlay nodes
  - This implements the support for new overlay protocols and network functionalities that address new application demands

Overlay Networks

- Software defined virtual connections:
  - Provide a **high customizability** for developers
  - Allow deploying **specialized protocols** that can provide new capabilities beyond what the Internet supports
  - Keep the same **scalable design** that has made the Internet so successful

→ **Cost-effective approach** to create a **customized network infrastructure**

- Although multiple layers of software and processing can increase **performance overhead** and make the network **more complicated**
  - Process of encapsulating and de-encapsulating packets can demand a significant amount of computing power
  - Physical network is not able to adjust automatically to changes in virtual networks

References

- Baker FJ (2009) Core Protocols in the Internet Protocol Suite. IEFT.
- Berners-Lee T, Bray T, Connolly D, Cotton P, Fielding R, Jeckle M, Lilley C, Mendelsohn N, Orchard D, Walsh N, Williams S (2004) Architecture of the World Wide Web, Volume One. W3C.
- Chandramouli R, Rose S (2006) Challenges in securing the domain name system. IEEE Computer Society 4 (1):84-87.
- Cisco (2018) Cisco Visual Networking Index: Forecast and Trends, 2017–2022.
- Grance T, Hash J, Peck S, Smith J, Korow-Diks K (2008) Security Guide for Interconnecting Information Technology Systems. NIST.
- IETF (2003) A Model for Content Internetworking (CDI).
- Jordana J (2002) Governing Telecommunications and the New Information Society in Europe. Edward Elgar Publishing, Incorporated.
- Mansfield KC, Antonakos JL (2009) Computer Networking for LANS to WANS: Hardware, Software and Security. Cengage Learning.
- Stevenson A, Waite M (2011) Concise Oxford English Dictionary. OUP Oxford.

Questions

Questions

1. How can a computer network be classified?
2. What are the differences between the Internet and the WWW?
3. Is the Internet a private or public network and why?
4. How are tier 1 and tier 2 ISP different?
5. How does IP-Routing work?
6. How are domain names structured?
7. What are the individual steps of the DNS look-up process?
8. What are the main benefits of a content delivery network?
9. What are the functions of the three SDN layers?
10. How does an overlay network work?

Further Reading

- Berners-Lee T, Bray T, Connolly D, Cotton P, Fielding R, Jeckle M, Lilley C, Mendelsohn N, Orchard D, Walsh N, Williams S (2004) Architecture of the world wide web, vol 1 W3C. https://www.w3.org/TR/webarch/. Accessed 30 April 2021
- Comer D (2015) Computer networks and internets, 6th edn. Pearson Education, London Hunt C (2002) TCP/IP network administration, 3rd edn. O'Reilly Media, Sebastopol, CA
- Mansfield KC, Antonakos JL (2009) Computer networking for LANS to WANS: hardware, software and security. Cengage Learning, Boston, MA
- Singh MP (2005) The practical handbook of internet computing. CRC Press, Boca Raton, Fl
- Tanenbaum AS, Van Steen M (2017) Distributed systems: principles and paradigms, 2nd edn. Prentice-Hall, Upper Saddle River, NJ