

Introduction to Networking (NSWI141)

Libor Forst, SISAL MFF UK

- Essential facts concerning communications
- Layered network model (OSI vs. TCP/IP, addressing, multiplexing, ...)
- Application layer (DNS, FTP, email, web, VoIP, ...)
- Transport layer
- Network layer (IPv4, IPv6, routing, firewalls, ...)
- Data link and physical layer (switch vs. repeater, Ethernet, Wi-Fi, cabling, ...)

Literature

- D. E. Comer, D. L. Stevens: Internetworking With TCP/IP; Prentice Hall 1991
- A. S. Tanenbaum: Computer Networks; Prentice Hall 2003
- C. Hunt: TCP/IP Network Administration; O'Reilly & Associates 1992

- internet resources

- Request For Comment (RFC)

- <http://www.warriorsofthe.net>

The “classic” books are listed here mostly for historical reasons. Currently, you can find many others in book stores and libraries if you prefer paper over a screen.

You can find some additional information on our topics on the Internet. Generally, information from the Internet is not absolutely correct in all cases. However, in this course we will not go into great depth, so the risk of finding wrong information is very low.

On the contrary, very precise information about most of the protocols we will study can be found in the official standards published (on the Internet) as so-called RFCs. However, it is a bit more detailed than is needed for our course, so usually only an Introduction chapter is enough for providing a sufficient overview of the topic.

For absolute newcomers, the Warriors of the Net video can be used as “a trailer” for this course, although it is a bit older and in some aspects simplifies things too much or formulates slightly inaccurately.

General attributes of communication

- Identification
 - actors must „find“ themselves (phone numbers) and introduce each other
- Method
 - e.g.: a deaf man at a counter tries sign language, the officer doesn't understand and suggests written form of communication
- Language
 - both sides must agree on a common language
- Speed
 - both sides must agree on a communication speed
- Process
 - requests, answers, confirmations

Introduction to Networking (2022) SISAL 3

We start by studying common ways of communication between people in real life, trying to find attributes similar to those used in network communication although you don't feel them as strict rules in real life.

Identification. In face-to-face communication you don't need special means for identification, your eyes and brain do the job naturally. The only exception is talking with an unknown person when an introduction is needed. In other communication ways you use some identification you are already used to using – a postal address, a phone number etc.

Method. If both sides are able to use all senses, there is typically no need to agree on a method. Of course, there may be exceptions as well. For instance if you use some signals, you must agree what they mean. If you are limited in using senses, an agreement on a communication method is necessary.

Language. In an international environment you need to agree on a language. In fact, a single common language is not always needed in this case, with natural languages you can also reach a reasonable level of understanding using two different languages, namely in the case of languages from the same family, e.g. Slavic languages.

Speed. Sometimes, speaking speed can cause comprehension problems, especially between native and non-native speakers. You will definitely feel a difference in understanding between a radio sports commentator and a teacher. In this case, again, some informal agreement on speed is needed ("slow down, please").

Process. In a normal situation, you will typically need no special strict rules. Everyone knows that in a conversation they have to address and greet their partner and say goodbye at the end. And if you forget something, it probably does not cause much trouble. However, when you are invited to visit the President, you will need to learn and obey a complicated protocol.

If we move to network communication, all these attributes will become much more formal and important; most differences can cause fatal misunderstandings while some can be accepted. But it is good to know that all of them are not something new; all of them come from real life, they have only been adapted and formalized.

Types of communication

- Ordinary human communications
 - voice, signals, writing
 - loose intuitive rules
- Telecommunication
 - complex technology with embedded rules
 - entire network (incl. end devices) is under centralized control
- Computer network
 - rules are open and freely accessible
 - most of logic is moved to end devices
 - network controls just the transmission
- Converged network
 - joins telecomm. and computer worlds (price, effectiveness...)
 - better is convergence based on computer network

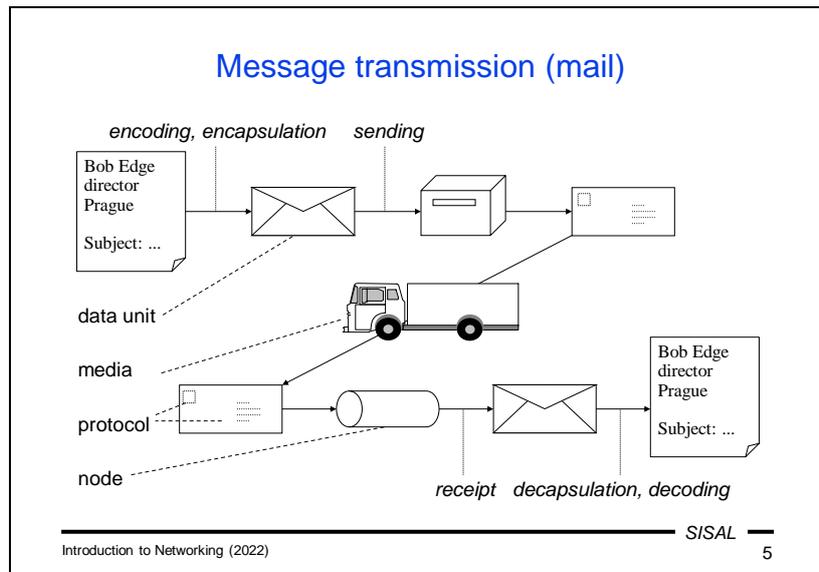
Introduction to Networking (2022) SISAL 4

Until now, we have talked about common forms of interpersonal communication and we have come to the conclusion that very loose and intuitive rules are used here.

The first global technology-based means of communication was telephony. From the beginning it was built as a private business so all nodes were under centralized control based on complex proprietary rules embedded into the transport devices. The end device (telephone) was only a simple interface for approaching the infrastructure containing all the application and transport logic.

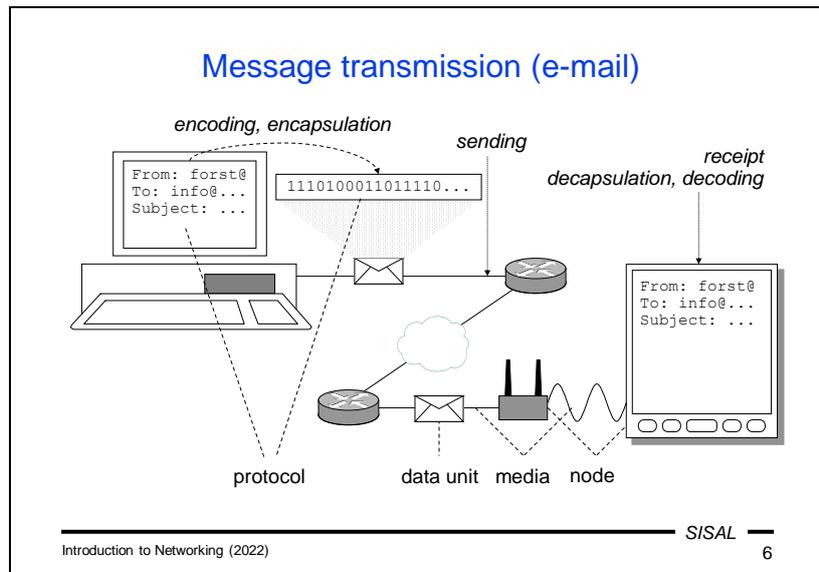
Computer networks were designed on a different principle: the application logic is concentrated in the end devices and the network infrastructure controls only the transmission logic. That's why it was necessary to design the system as open and publish the rules so that each application implementer can behave according to them.

When studying the communication, we must necessarily ask the question of how to connect both types of networks. Over the history, there have been several modes of such interconnection. Originally, computer networks exploited the telecommunication infrastructure (leased lines). Later, they built their own cabling systems and the situation became inverted: nowadays, telephony is very often transmitted over a computer network (Voice over IP). However, the boom of smartphones with mobile data has once again reversed the situation in many cases – the network communication is again transmitted over telephone channels.



Let's look at the process of transmitting a message by (classic) mail. You take a sheet of paper with the message, put it in an envelope, write an address, attach a stamp and throw it into the mailbox. Then, the post organization handles the message using its infrastructure to deliver it to the recipient's mailbox. Upon delivery, the recipient opens the envelope and takes out the message.

Now, if we want to describe the process by means closer to network communication, we can start to use technical terms that are not commonly used but fully understandable. You take a sheet of paper with the message [data unit], put it in the envelope [encapsulation], write an address [target node address], attach a stamp [follow the mail protocol] and throw it into the mailbox [sending]. Then, the post organization handles the message using its infrastructure [intermediate nodes connected by various media] to deliver it to the recipient's mailbox [target node]. Upon delivery, the recipient opens the envelope [decapsulation] and takes out the message [message delivered].



Using very similar words we can describe the message transmission process in the case of electronic mail. We also need to use encoding and encapsulation according to a proper protocol to create a data unit that has to be sent to the local network. Then the networking infrastructure handles the message over various intermediate nodes and various channels to be delivered to the target local network and through it to the end node (your personal computer, smartphone etc.) where it is decapsulated, decoded and displayed on a screen.

We can see here the same actors, the same roles and the same tasks like in the classic mail, just converted to a technical platform. The basic principle of functioning is in both cases the same.

Requirements - fault tolerance

- circuit switching: faster, fluent, however failures disrupt whole connection
- packet switching: packets use various ways => transport time may differ, but network can overcome node failures

Introduction to Networking (2022) SISAL 7

At the inception of the idea to use electronic means for message transmitting, there were several basic requirements. Some of them remained till nowadays, some of them have changed somewhat over time.

A crucial requirement was fault tolerance. In fact, the army (the US Department of Defense) was behind the whole idea. They were solving the major problem of telephone network outages in the event of an enemy attack. The telecommunication network uses so-called *circuit switching*. This means that when you call someone, the network will find a sequence of nodes (a circuit) needed to connect your device and the end device that you are calling. Once this circuit is established, all audio data follows this path. If the enemy breaks a node in the circuit, the circuit is lost.

The solution is pretty easy. We divide the data into smaller blocks so-called *packets* and let each packet to find its own way to the target node. If a node is broken, the packet will find an alternate way. It will be delayed a bit, but it will be delivered. Of course, nowadays the reason for node failure may not only be an enemy (hacker) attack, but also a power failure, misconfiguration etc. So risks still exist and therefore this principle is used by current networks for data delivery.

When comparing these methods, we must say that circuit switching is faster but unreliable, while packet switching is slower but fault-tolerant.

Requirements – security I

- Original motivation:
 - physical security of transmission

- In the early days of networks, there was little risk of network attack other than physical destruction

- Old technologies were naive:
 - open communication (tapping possible)
 - full confidence in partner identity
 - content trust

Introduction to Networking (2022)SISAL

8

Fault tolerance is closely related to the more general concept of security, specifically physical security, i.e. the security of the physical infrastructure (nodes and cables). Surprisingly, the other side of the coin – logical (data) security was not taken into account when designing the networking model. The reason is probably that the size and price of computers at the time, along with the complexity of programming, was a major obstacle to an attacker being able to get close enough with his device to attack data.

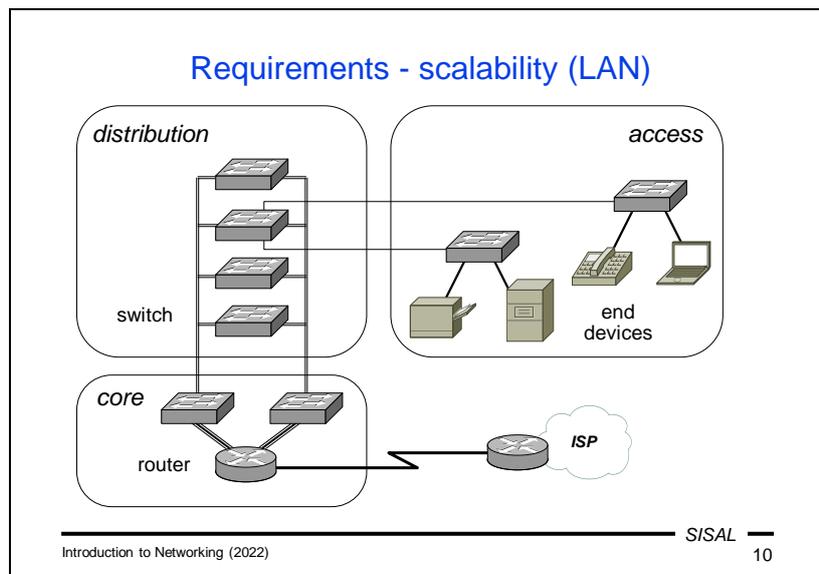
Currently, the situation is the opposite – it is very easy to have a computer capable of connecting to the network and it is quite easy to infect a target computer by attacker software so that you can see or even control the transmitted data.

All old protocols have the same problem – they do not use any encryption and trust both the “sender” of the data and the content of the data. Unfortunately, you can see the same behavior now among many people who trust everything they see in the mail or on the web.

Requirements – security II

- Security risks:
 - infrastructure (physical) security
 - data (logical) security
 - DoS (Denial of Service)
- Infrastructure security:
 - access restrictions
 - backup power sources, air conditioning...
 - backup servers, connectivity...
- Data security:
 - user authentication, access rights control
 - host authentication (servers, clients)
 - data inspection (application proxy, antivirus, antispam, ...)
 - cryptography (ciphers, encryption, subscription)

Of course, the requirement for physical security has not disappeared even today. In addition to the risks resulting from unauthorized physical access, technical risks such as power outages, connections, etc., were added to at least the same severity. Over time, data security requirements (i.e. prevention of unauthorized access to data or unauthorized data manipulation) have gained much more weight. With the progressive development of connectivity and availability of IT equipment, the risk of DoS (Denial of Service) attacks has also increased significantly. DoS is an attack where an attacker attacks a data source with massive communication, so that it cannot communicate with regular users. In addition, such an attack is usually strengthened in such a way that the attacker takes advantage of some errors/weaknesses in the implementation of a certain service and forces foreign servers to unwittingly generate increased traffic instead of him (the so-called DDoS - Distributed DoS).



One very important requirement is scalability. This means that the system must be designed to make adding a new computer or network easy, without having to make changes in remote locations or parts of the network. Scalability leads to logically similar but technically different approaches when adding a host to a (local) network and adding a network to the Internet.

A recommended approach to design a local network is dividing the infrastructure into three parts:

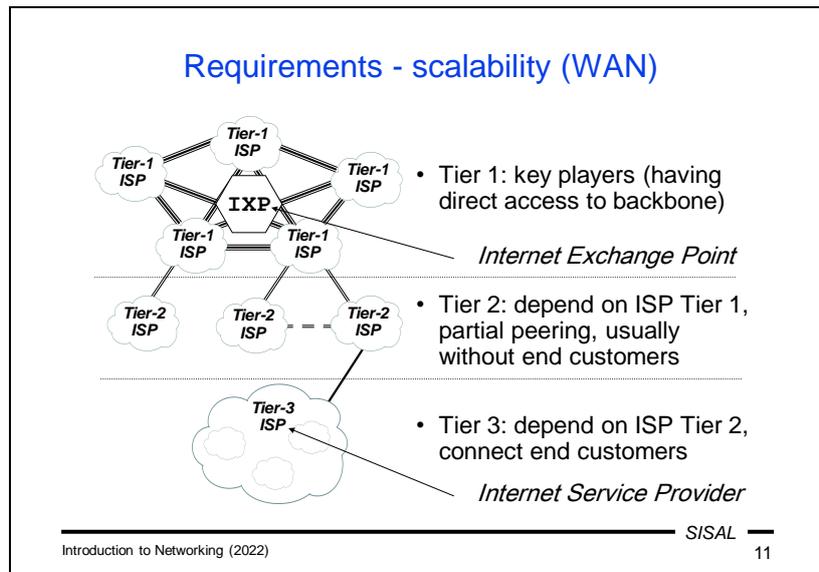
The **core** layer is the main part of the network, where the backbone of the network is terminated and the network is connected to the infrastructure of an Internet Service Provider, which mediates the Internet connection for our network. Typically, the core layer nodes are located in a special room with air conditioning and an uninterruptible power supply. The key device there is a *router*, a node interconnecting different networks. This router is an entry point of our network, connected to another router on the ISP's side. Another important device is a *switch*, a node interconnecting other switches and end devices in the network. The core layer of the local network usually contains a few main switches connecting the perimeter router to the rest of the local network.

The **distribution** layer is a backbone part of the network, responsible for distributing of the connectivity to all parts of a building or campus. Sometimes, it is also called a *vertical* layer since it is very often located in a vertical well going through all floors of a building.

The **access** layer is the last part of the network, allowing access to network service to end devices, servers, printers, personal computers, phones etc. Sometimes, it is

also called a *horizontal* layer. This layer is terminated as close as possible to where end users need to connect.

This approach to local network design guarantees easy network expansion by adding new hosts. Of course, it can happen that we reach the maximum capacity of some part of the access layer, but in this case the further network expansion again affects only particular parts of the distribution layer.



A similar concept of a three-layer model is used also for the world behind a perimeter of a local network.

At the top of the hierarchy, there is a layer called **Tier 1**. In this layer, the key players of the Internet are included, i.e. companies having direct access to the Internet backbone interconnecting all continents. Examples: AT&T, Deutsche Telekom, British Telecom.

The bottom layer of the tree is called **Tier 3**. There are included ISPs who connect end customers – typically companies, organizations, households etc. connecting their local networks to the Internet.

Somewhere in between these two extremes lies the **Tier 2** layer with companies without direct access to the backbone, whose customers are other ISPs. Typical examples are regional or national ISPs.

Requirements - quality of service (I)

- Network transmission parameters:
 - latency, delay
 - evenness of delivery (*jitter*, variance of delay)
 - data loss
 - bandwidth („speed“)

- Various applications have various requirements:
 - multimedia applications: low jitter
 - data transfer: low data loss
 - ...

Introduction to Networking (2022)SISAL

12

The last requirement is a sufficient quality of services provided by the network to individual applications. However, this requirement does not lead to a single correct approach since for various applications different network transmission parameters are important.

The main transmission parameters we can investigate are:

- **Latency** is in fact the communication delay, i.e. the amount of time from when a piece of data is sent to the network to when it is delivered to the destination.
- **Jitter** is the variance of the delay; this value expresses how regularly the data is delivered, i.e. whether the volume of received data fluctuates.
- **Data loss** means how often a data packet is not delivered; high data loss means either loss of information during the transmission, or the need of repeated sending of lost data packets.
- **Bandwidth** is often called “speed” which is a bit misleading since the speed of transmission of the signals over a given media is always the same (no ISP can use faster light for transmission over an optical fiber); in fact, it means how much data can be encoded and transmitted using particular physical signals, not the actual speed of the signals.

When we look at these parameters, it’s easy to understand that the insufficiency of some of them brings serious problems for various applications. For example, multimedia applications use special encodings which are robust against data loss, so it is not a problem to some extent. In contrast, high latency can be critical when using such applications in real time (e.g. phone calls), but it does not have much impact when watching a movie. But in any case, if the jitter is high, no multimedia application will work well. On the other hand, high data loss is critical for applications based on

entire data transmission, like web, email etc. Here, it causes channel congestion due to massive data resubmission.

Requirements - quality of service (II)

- Goal:
 - guarantee dedicated throughput for particular traffic types
 - guarantee faster delivery of priority messages

- Implementation:
 - data is classified by QoS tag
 - *guaranteed service* strategy: dedicated part of channel
 - quality guaranteed, wasting of channel capacity
 - *best effort* strategy: priority queues
 - effective media usage, quality not guaranteed

Introduction to Networking (2022)SISAL

13

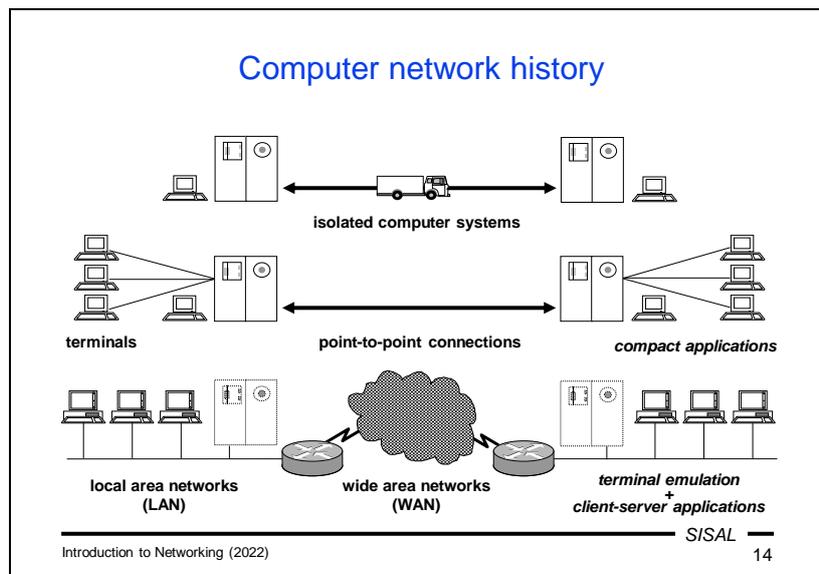
In general, we would like to achieve two basic goals:

- For some traffic types, we want to dedicate a bandwidth that is sufficient for a specific application to work well.
- For some traffic types (or at least for some types of messages) we need to guarantee fast enough delivery – e.g. during a massive data transfer by one application we still need to deliver urgent messages quickly.

Here are the current methods used to achieve these goals:

- The primary condition is how to recognize different data transfer types. For this purpose, the sender marks packets with a special value called the *QoS tag*. Of course, this approach only works if all applications behave fairly. If an application starts cheating and marks all data with a high-priority tag, this system crashes.
- For applications for which we need to guarantee throughput, we can separate a part of the channel bandwidth and dedicate it to the application. This approach works well if we have enough bandwidth for all the necessary traffic. Otherwise, we have to reduce or cancel the bandwidth reserved for certain traffic. The problem with this approach is that if a particular application does not currently use its dedicated bandwidth, it is simply unused, wasting the channel transmission capacity.
- Ensuring the appropriate speed of messages according to their declared priority can be solved by the so-called *best effort* strategy using priority queues. A network node has different queues for messages with different priorities and selects messages for sending from different queues accordingly. The advantage is that there is no waste of channel capacity, but delivery speed cannot be guaranteed under heavy load. The problem is mainly on the input side of the node - if a packet arrives at the node, the node cannot recognize the packet type or priority classification until it reads a sufficient part of the packet. During this time, the node

must deal with the packet, although it may possibly have a low priority, while other higher/priority packets may wait to be received and processed.



What was the situation at the inception of the idea of a computer network? There were only a few isolated computers with very limited means of interaction both between a computer and users and among computers with each other. A mainframe computer usually had a single alphanumeric operator console from which all jobs were controlled. Input and output devices were limited as well – e.g. punch card readers were used to enter a larger amount of data. If it was necessary to transfer data between computers, you had to punch the data into a set of punch cards, transfer them to the other computer and read them in there.

Later, computers could serve many more users, so it was necessary to add some way to allow users to access a shared computer. The solution was terminal networks. A *terminal* is a simple device with only a keyboard and monitor that can transmit keystrokes from a user to a computer and display the response on the user's screen. For communication between two computers, a proprietary point-to-point communication connection was built (usually via telephone leased lines) without any more complex structure. The applications used at that time were mostly still the same compact applications as in previous era.

In the last phase of development, two important changes took place. Random point-to-point connections were converted to a global structured network, a so-called *Wide Area Network*. On the local side, terminals were replaced by personal workstations and were connected each other (and connected to a main site-local computing facility if it remained) by a replacement of the old terminal network – a *Local Area Network*.

One of the consequences of this change was also the introduction of a new model of application, the so-called *client-server* model. It's based on the principle that a part of the computing power is transferred to the front-end computer, so an application has

two parts, a *client* running on your computer and a *server* running somewhere on the network. These two parts communicate with each other, transmit user requests from the client to the server and mediate the responses to the user. Thus, the terms client and server in their original meaning do not refer to computers, but to software.

Basic network types

- Local Area Network (LAN)
 - resources sharing (file- and database-servers, printers,...)
 - shorter distances (building, campus), minor delay
 - proprietary networks, centralized control
- Wide Area Network (WAN)
 - remote access, end-to-end communication
 - large distances, notable delay
 - multiple owners, distributed control
- Nowadays:
 - differences fade away (most important is possession)
 - interlayers occur (MAN)
- Classification is not technical (no definition), but logical

Introduction to Networking (2022) SISAL 15

After this trip into history, we can look at the terminology used for types of networks according to their scope.

Local Area Networks are descendants of terminal networks, which means that their main purpose is to share resources between nearby computers (we can also consider an Internet connection as one of the shared resources). They are usually geographically limited by the boundaries of a building or campus whose owner builds and manages the network primarily for its employees, students, customers, etc.

A **Wide Area Network** is a global network that has replaced point-to-point connections between computers or computer centers that were built on demand. Its main purpose is interpersonal and inter-computer communication and data transfer. It is built and managed by many companies and used not only by people related to them.

In between these two ends of a wide range of networks, there are a few more categories without an entirely precise meaning, e.g. Metropolitan Area Networks.

In the past, there were also fairly well-defined technical differences between a LAN and a WAN, such as the technology used, bandwidth etc. Over time, these differences have disappeared and currently the main difference is in a relation between equipment owner and users. Unlike WAN, LANs are usually private. **No definition like “A network is a LAN is when...” exists.**

Public and private networks

- Most LAN are private (user is the owner)
- Most non-LAN are public (user is not owner)

virtual private network (VPN)

- VPN motivation: security, expenses
- Typical use of VPN: affiliates interconnections, connection of (mobile) end users

Introduction to Networking (2022) SISAL 16

As already mentioned, LANs are in most cases private, while WANs are mostly public. However, this poses a problem for the network designer if he needs to expand a private LAN geographically over a large area, for example, between affiliates in two distant cities. Connecting such sites with private cabling is not realistic and, by using a public network, we lose privacy.

The solution is called a **Virtual Private Network (VPN)**. The basic idea is that on the perimeter of each part of a LAN there is a device that is connected to its counterpart via a public network using a so-called *VPN tunnel*. All the traffic directed to the other affiliate is encrypted at the local end of the tunnel, sent over the public network, then decrypted at the other end of the tunnel and delivered to the second LAN. For computers in both LANs, the whole mechanism is transparent and they do not need to take care of it.

Another modification of this principle is when the whole side of a tunnel is replaced by a piece of software running on a personal computer (notebook). Then this computer seems to be connected (through the tunnel) as a normal node in the private network to which it belongs.

Note: At the end of the semester we will learn another term that sounds very similar, Virtual Local Area Network (VLAN), but its meaning is totally different, so do not confuse them.

Internet history

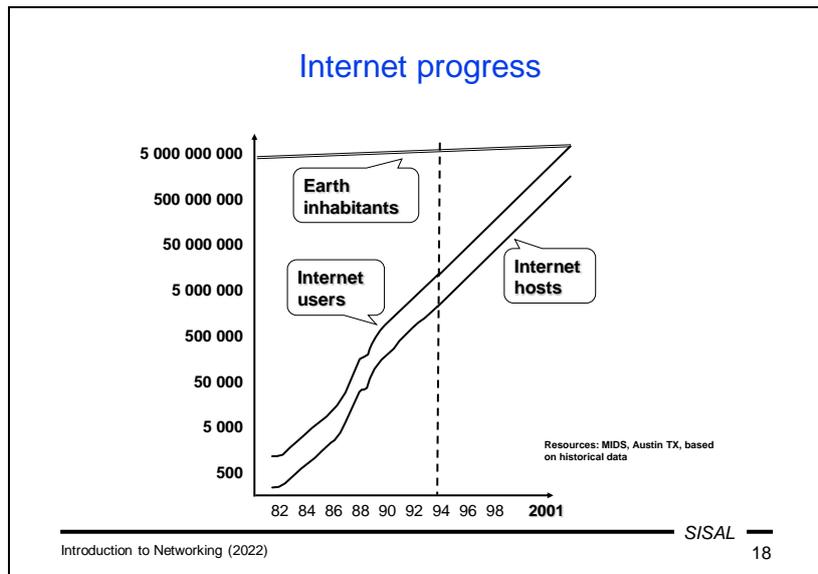
- beg. of the 60s - „packet switching“ concept
- the 60s - US DoD supports „packet switching“ concept for its resistance against a physical attack
- 1969 - ARPANET - paid by Defense Advanced Research Project Agency, managed by academic institutions, point-to-point leased lines
- 1974 - term „Internet“ (abbr. of „internetworking“) used in RFC 675 defining TCP
- 1977 - first network bound to ARPANET backbone
- 1983 - TCP/IP replacing NCP in ARPANET
- half of the 80s - TCP/IP included into BSD UNIX

In the early 60s the concept of packet switching originated, supported by the US Department of Defense. It created a specialized agency Advanced Research Project Agency (ARPA) where several academic institutions tried to implement this idea and create a network based on a structure of point-to-point connections over leased telephone lines. This network was created in 1969 and called the ARPANET.

Originally it interconnected only separate computers and in 1977, the first network was connected to the ARPANET backbone. In fact, this moment actually gave another meaning to the term “Internet” which appeared in 1974 – an interconnection of networks.

Currently, the leading networking technology is TCP/IP; it replaced the previous one, Network Control Program (NCP), in the ARPANET in 1983.

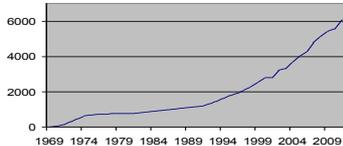
The last key development milestone came in the mid-80s when developers of the Berkeley System Distribution (BSD) clone of the UNIX operating system understood the importance of TCP/IP and added support for it directly to the core of their UNIX. At that time, all users of the system were given the means to connect to the Internet.



Just as an interesting fact that proves how dramatically this branch was developing, we can look at a chart published in 1993 by the Internet Society. They used an extrapolation of the network development in the preceding four years and the predicted result was that in 2001, the number of people using the Internet (people, not user accounts) would exceed the population of the Earth...

Request for Comments (RFC)

- Means of Internet „standardization“
- RFC 1 published Apr 7, 1969



Year	Number of RFCs
1969	0
1974	~500
1979	~1000
1984	~1500
1989	~2000
1994	~2500
1999	~3500
2004	~4500
2009	~6000

- Freely accessible (<http://www.ietf.org/rfc.html>)
- Various nature: standards, information, best-practice
- Drafts are sent to and judged by IAB ⇒ IETF, IRTF ⇒ WG
- Document texts are fixed, upgrades obtain new number (SMTP: 772, 780, 788, 821, 2821, 5321)
- Current status can be found in the index file
- Recommendations are widely violated

SISAL

Introduction to Networking (2022)

19

Currently, Requests for Comments represent a means of standardization on the Internet, although their name does not suggest this.

Originally they were used for presenting ideas and discussing them together with publishing various data. That's the reason for the name. As an example of what was presented by them in the prehistory of the Internet, we can mention the fact that e.g. changes of computer addresses were published via RFCs. This practice of relatively free contents ended in the 70s and as you can see in the chart, the increase in the number of RFCs decreased rapidly.

Nowadays, their publishing is controlled by the Internet Advisory Board through two commissions, Internet Research Task Force and Internet Engineering Task Force, and several working groups. When an author submits a draft of a new protocol, the relevant working group judges it and, if it considers it useful, the document receives a number and is **published** "for commenting" (as the original meaning of "RFC" says). The important fact is that the text of the document **never changes** (except for typos and errata). You don't need to search the internet for the newest edition of an RFC. All servers have the same version. When there are a sufficient number of changes, the document is reissued with **a new number**. RFC number changes can be tracked in an index file called `rfc-index.txt`. The old RFC can be marked either "Obsoleted", or just "Updated", in which case both RFCs are valid.

Adherence to the protocol rules defined in the RFC is very important for all implementers. However, some of the rules are a bit restrictive and many clients and servers violate them. The mutual understanding of communicating parties then depends on the extent of the violation. The general recommendation for

implementers is: be as tolerant as possible on the receiving side and be as conservative as possible on the sending side!

Summary 1

- What are the advantages and disadvantages of packet switching?
- How have network protocols been affected by the fact that the idea of networking was initiated by the army to increase the security of communication?
- What is the purpose of the network scalability requirement?
- How do the demands of e-mail and IP telephony differ regarding the transmission parameters of the network?
- What is the definition of a LAN?
- What is the essence of a VPN?