What is a computer network?

- A distributed physical device.
- A directed graph.
- A facility for exchanging data.
- A facility for digital multimedia exchanges (voice and video; perhaps more).
- To most people: the Internet.

This course is about networking using the Internet and other networks that are similar to the Internet.

Basic units

In the world of networking, the units follow SI conventions, i.e. k = 10^3 (not 2^{10}), M = 10^6 (not 2^{20}), etc.

ms, μs, ns, ps units of time.

$$1s = 10^3 ms = 10^6 \mu s = 10^9 ns = 10^{12} ps.$$

kHz, MHz units of frequency (defined as multiples of s^{-1}).

kb/s, Mb/s, Gb/s, Tb/s are units of bandwidth (“capacity”) of a network connection. Frequently called “speed” (which is nonsense: most networks operate at the same speed). Note that **b** stands for **bit** not **byte** and that bps = b/s.

Make sure not to use simplistic assumptions about the relationship between bandwidth and frequency: a signal send at 2.4 GHz normally does not carry 2.4 Gb/s.

Propagation speed of a signal

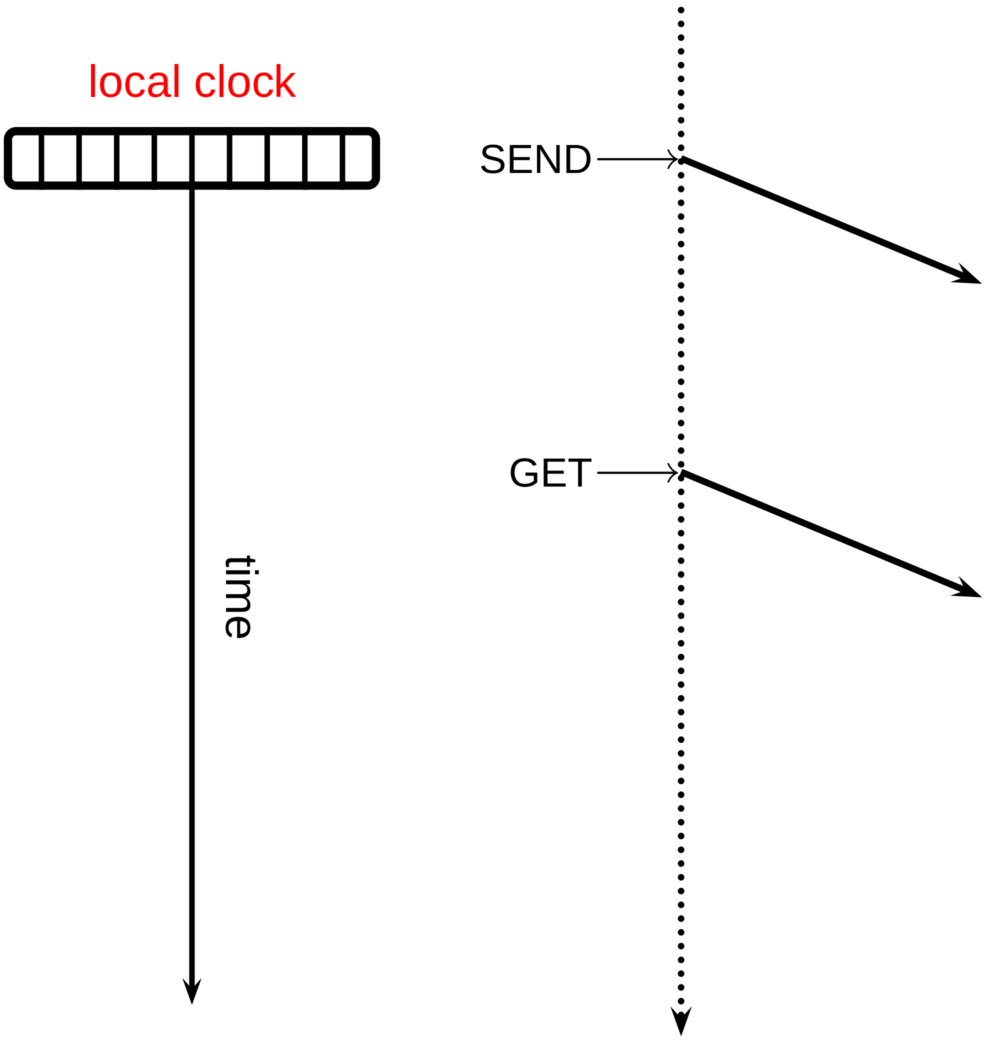
In communication networks of today, the main transmission media are electricity (wire) and light (fibre or air).

The speed of light is not constant, but varies depending on the medium. The exact values are hard to use, so two handy approximations were adopted: 2×10^8 m/s for glass and 3×10^8 m/s for air and vacuum (note that a more exact value for vacuum is 299,792,458 m/s).

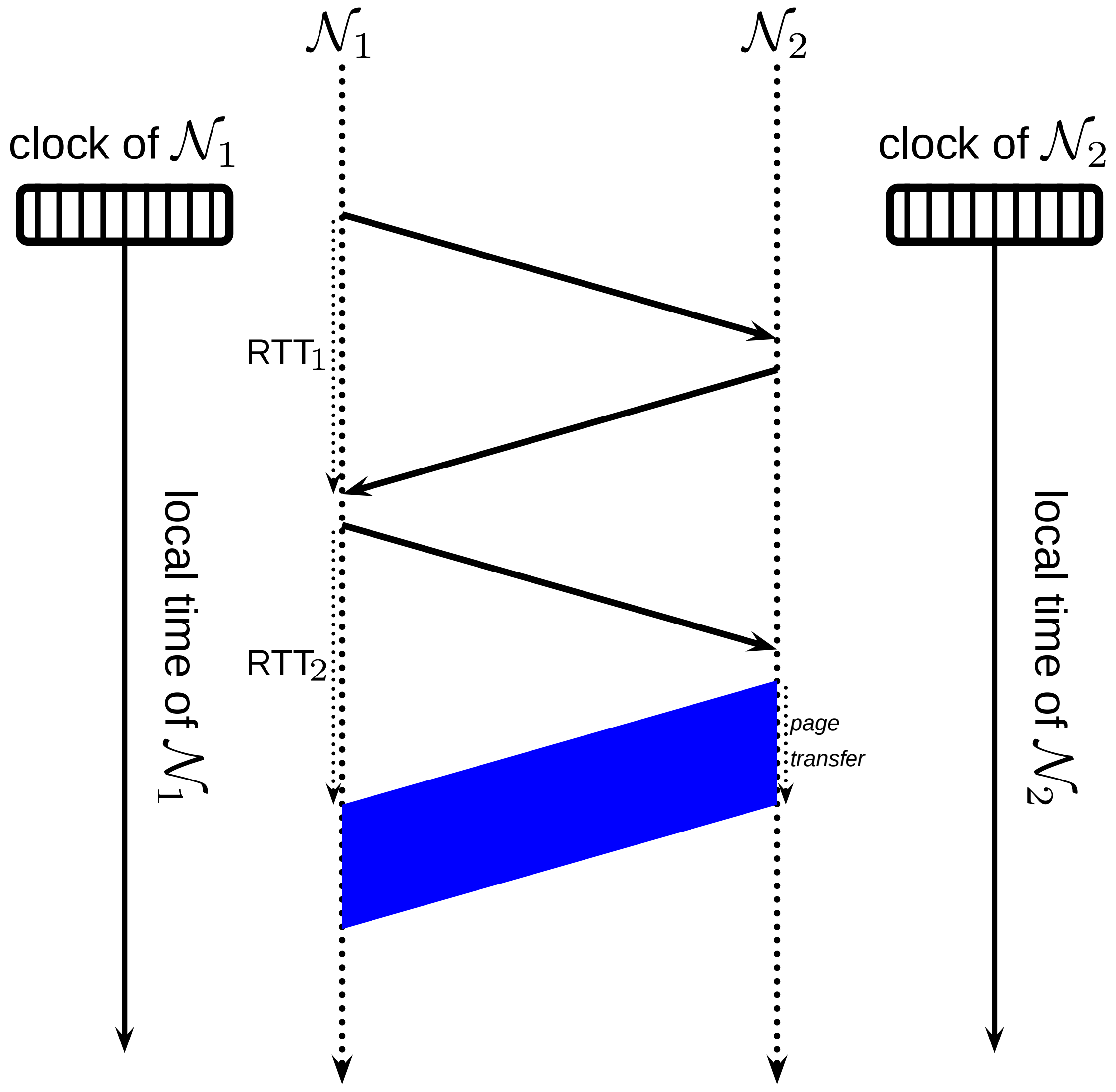
Electrical signal propagates at the speed of 1.92×10^8 m/s in **CAT5** UTP copper cable and 2.31×10^8 m/s in a **CAT7 STP** cable. Unless it really matters, a “one size fits nobody” approximation of 2×10^8 m/s can be used.

Reading communication diagrams

The timing of events in a network node is captured by timing diagrams like this:



Time flows downwards and always means the value of the **local** clock.


$\mathcal{N}_1$
$\mathcal{N}_2$
clock of $\mathcal{N}_1$
clock of $\mathcal{N}_2$
RTT$_1$
RTT$_2$
page
transfer
local time of $\mathcal{N}_1$
local time of $\mathcal{N}_2$

Composition of a network

Seen from the outside world, a network is made of two parts: the **edge** and the **network cloud** (also called **core**).

The edge is made of two sets of entities:

Users which can be people or devices (sensors, refrigerators, security cameras, etc.).

Hosts also called *end systems* are the computers and computer–based devices that connect users to the the network cloud (the inside of the network). The number of Internet hosts exceeds 10^9, but not all hosts ever accessed the Internet.

Composition of a network

The network cloud is made of:

Access points (base stations, switches, modems, cell–phone towers) which connect the hosts to the routing part of the network.

Routers and switches, gateways, etc. form the routing part of the network core. They are responsible for passing packets from hosts to hosts.

Link is a communication medium connecting two entities.

Circuit is a collection of links connecting two hosts.

Packet is the name of a block of data that is sent between two entities either over a link or a circuit.

Access points are doors into the core; routers are transfer points inside the core.

Switches have a dual nature because hosts are connected to them, but other switches may as well (a switch–switch link is inside the core; a host–switch link is outside).

Routing

This topic will be explained when the **IP** protocol is discussed.

Briefly, it is the process of choosing the next link (or several next links) which will be used to transfer a packet from its current location towards its destination. A sequence of routing decisions creates a circuit.

A simple routing decision (choice of the next link to use) results in a single **hop**; a circuit can thus be seen a sequence of hops.

Major abbreviations

Several cryptic abbreviations are used by the networking community:

IANA (Internet Assigned Numbers Authority) is the central coordinator acting as a clearinghouse for assigning numbers.

IETF (Internet Engineering Task Force) is the manager of the Internet protocols and thus naming conventions (IP addressing), etc .

RFC (Request for Comment) is a document containing a version of an Internet standard. The RFC numbers increase with time, so a larger number supersedes a smaller one (RFC793 $\Longrightarrow$ RFC1700 $\Longrightarrow$ RFC3232, etc.).

Addressing

All the network entities have network addresses to identify them by other entities. There are several types of addresses, each related to a particular software layer in the network:

Logical addresses such as

www.cis.uoguelph.ca

used by applications (including humans). They should not be confused with URLs although they have things in common.

IP addresses are numeric counterparts of logical addresses and look like this:

131.104.48.133

Physical (MAC) addresses representing physical **N**etwork **I**nterface **C**ards (NICs). Their format, expressed in hexadecimal, looks like this:

1A:23:F9:CD:06:9B

Ports

A typical host is capable of running several activities (processes) concurrently. Therefore it is not enough to give the host's address to reach one's counterpart. Hence each host has an array of **port**s, each port tied to a particular activity (or idle if not tied to anything).

Ports are numbers from 1 up to a maximum of $2^{16} - 1$ (65,535). They are divided into three groups (RFC 3232):

Reserved ports known as **Well Known Ports** (see IANA:26–08–2008 or wikipedia). They have numbers from 1 to 1023 and are indeed reserved by IANA.

Registered ports numbering from 1024 to 49151. These ports are not to be used without registering their use with IANA.

Private ports numbering from 49152 to 65535 which are free to use.

Nobody pays any attention to the need to register a Registered Port. We may consider using a registered port 3210 assigned to:

flamenconetworks.com

and dubbed

flamenco-proxy

No problem will arise unless the actual *flamenconetworks* happens to be the node we want to communicate with (which seems impossible).

End–to–end connections

From the user's perspective, the network is made of hosts which come in two varieties:

Clients are hosts that request (and usually receive) services from other hosts. They join the network when they need it; otherwise, they are dormant.

Servers are hosts that provide services to clients. They are expected to be permanently active, listening for incoming requests. Note that any running computer with an operating system that supports **Well Known Ports** is a potential server because it can listen at least to ports 25 (SMTP) and 80 (HTTP) and many other ports.

Note that a host may be a server and a client simultaneously (e.g. conference calls, mirror database sites). The special case when two clients communicate with each other directly is called **peer–to–peer** or **P2P**.

Network size

LAN: local area network, defined loosely as a network without routing needs, messages being broadcast through a shared medium. Hub–based Ethernet and *switched Ethernet* (multiple Ethernet segments connected by a routing switch) make the definition a bit confusing.

MAN: metropolitan area network, in which point–to–point messages are not broadcast and routing is needed.

WAN: wide area network which is like a MAN but must be **scalable**. WANs use heterogeneous hardware and are connecting smaller networks, hence *internet* is a synonym for a WAN.

The Internet as opposed to *any* internet, is spelled with a capital I.

Bits

Information can be sent “as is” (“analog”) or encoded (“digital”). Synonyms “continuous” and “discrete” are often used,

We care very little about analog transmission nowadays, so it will be left out of this course. The **circuit switching** approach to handling analog transmission is also ignored (in this approach, a continuously reserved circuit is set up at the beginning of a session).

In a digital network, the **bit** is the atomic unit of transmission, even though some coding methods encode several bits as a unit (e.g. FDDI). Bits are typically grouped into units called **messages**, **packets** or **cells**.

Digital networks can use either circuit switching or packet switching as a method of sharing links. The Internet uses **packet switching** (or message switching) even though **virtual circuits** are most likely set up on top of the network.

Multiplexing and circuit–switching

When a network link is shared by many circuits, sharing is a necessity. In C–S networks, it is done in a preassigned manner using:

FDM (frequency–division multiplexing). Each circuit is assigned a frequency band within the range available in the link. The band is used continuously by the circuit.

TDM (time–division multiplexing). At regular intervals, the whole link is made available to each circuit for a fixed amount of time, called a **slot**.

Multiplexing in telephony

In a T1 1.544 Mb/s link sets 8 kb/s for control, with the remaining 1,536 Mb/s shared by 24 circuits, each circuit will get:

- One continuously available virtual link of 64 kbs when FDM is used.
- One slot of size 8 bits every $\frac{1}{8000}$ of a second when TDM is used[a].

Telephone networks used FDM until the 1960s when PCM using TDM was introduced.

[a] It could just as well be 1 slot of size slot of size $\frac{64}{24}$ kb every $\frac{1}{24}$ second. Etc.

Multiplexing and packet–switching

CDMA (code division multiple access. All circuits transmit simultaneously, with each circuit assigned a unique code which spreads the signal. Only a receiver possessing the same code can recover this signal; all others will hear it as noise. CDMA limits the number of circuits to the number of available codes, but does not waste bandwidth when moments of silence occur in some circuits. (See details) The method is used in wireless telephony (e.g. 3G).

Statistical multiplexing has no fixed ordering; packets are sent on a first–come–first–serve basis (with variants).

Unlike channel division multiplexing, statistical multiplexing does not **guarantee** an advertised bandwidth but gives a **best effort** bandwidth. CDMA is somewhere in between.

The bogey man: delay

Processing delay: time elapsed between the moment a packet is received and the moment it is forwarded to an output queue.

Queuing delay: time spent waiting in a queue before being transmitted.

Transmission delay: time elapsed between transmitting the first and the last bit of a packet.

Propagation delay: time elapsed between the moment a bit is transmitted and the moment it is received.

For details, see reference).

In multi–hop network, each delay occurs during every hop.