Information vs. Data

According to Russell Ackoff the content of the human mind can be classified into five categories, two of them being of interest here:

Data: symbols.

Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions.[a]

Knowledge

Understanding

Wisdom

[a] And many other questions, in my opinion.

Telecommunications

wikipedia says that

> **Telecommunication** *is the assisted transmission of signals over a distance for the purpose of communication.*

This broad definition includes any kind of communication method; the advent of computer networks resulted in the development of a specific area of telecommunication:

Digital Data Communication.

Rapid progress of digital techniques obsoleted most analogue telecommunications methods and all the significant telecommunication methods are becoming digital data signalling methods.

This course is devoted to digital telecommunication with particular emphasis on communicating **data**.

Audio and video issues will be covered sketchily.

Note that there is no consensus as to whether telecommunication or telecommunications is the right word.

Data communication

According to Forouzan, data communications is:

> *the exchange of data between two devices via some form of transmission medium such as a wire cable.*

This definition exposes the problems of trying to give rigid definitions to real–life phenomena: in reality, broadcast communication involves more than two devices; also, anything can be a medium as illustrated by quantum communication.

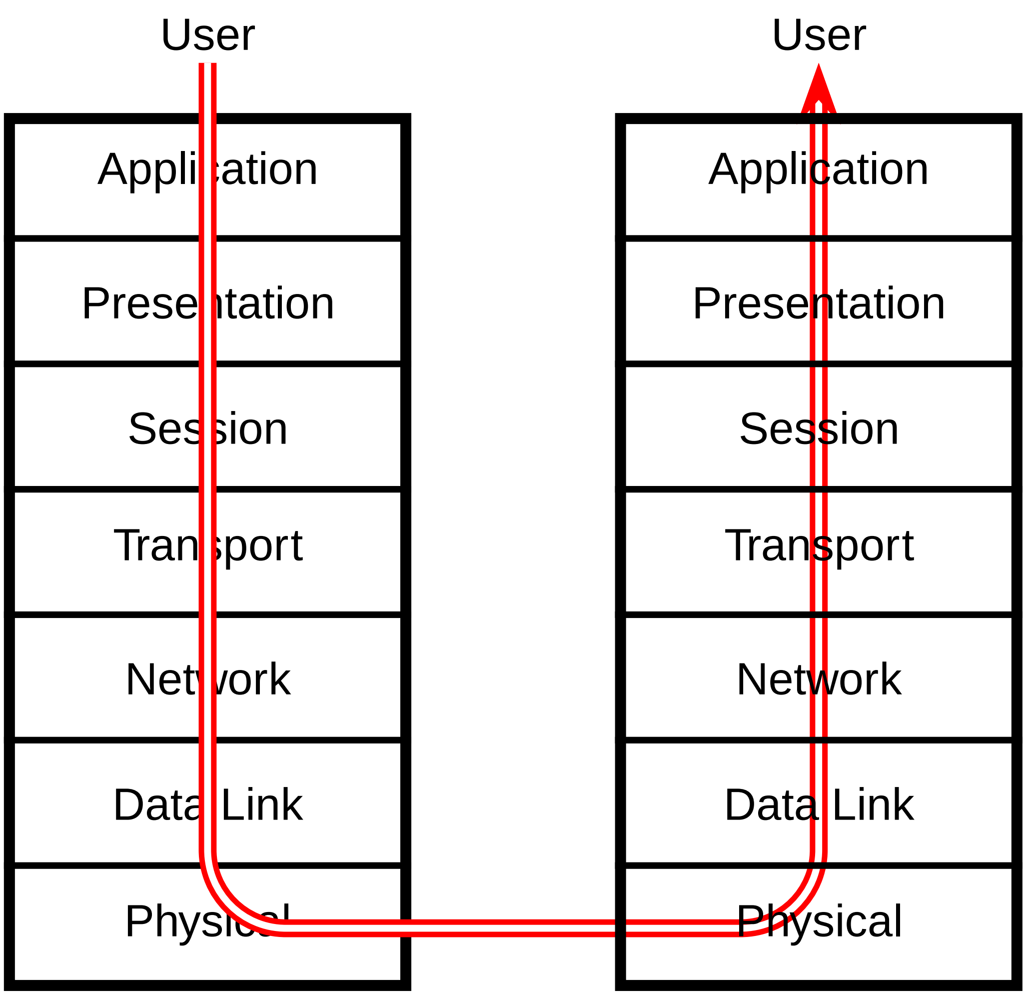
Computer networks

The primary carrier of digital communication is a computer network which a combination of software and hardware. The **OSI** model describes a network in an abstract way:

Application layer
Presentation layer
Session layer
Transport layer
Network layer
Data Link layer
Physical layer

A user message travels **down** the layers to the very bottom (the sender's telecommunication hardware).

The hardware transmits the message to the receiver.



When the message reaches the receiver's hardware, it starts moving **up** until it is handed to the receiving application.

This course

deals with the operation of the two lowest layers of the network hierarchy: the Physical Layer (**PL**) and the Data Link Layer (**DLL**).

These two layers are responsible for moving around raw data, i.e. blocks of bits that have no meaning associated with them.

The higher network layers gradually add a notion of meaning to these data blocks, ending with meaningful information seen by the receiving Application Layer entity. We do not deal with information in this course.

Physical Layer

This layer transmits individual bits on a single physical link, which may include repeaters. There is no error control at this level.

Conveniently, there is no definition of a link that could be of value. The concept will be explained intuitively in a moment.

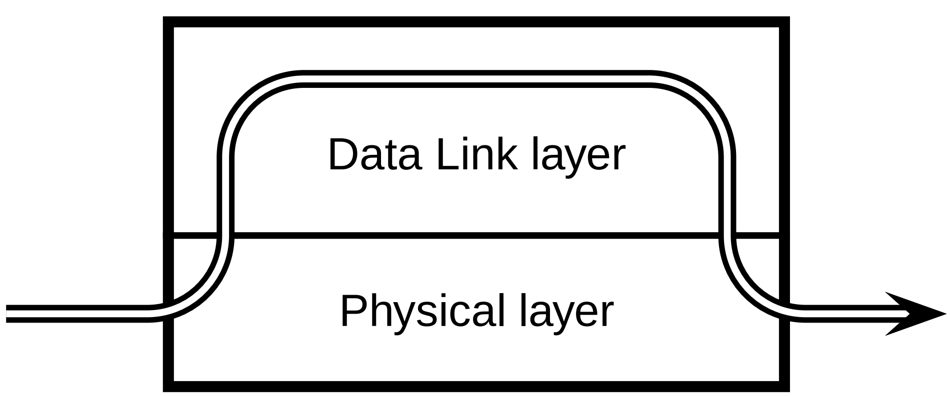
What constitutes a single link is product dependent.

(Data) Link Layer

This layer transmits packets one hop forward, along one physical link. Examples: *Ethernet* (CSMA/CD), *PPP*. Note that a single link may include bridges (e.g. multiple–segment Ethernet).

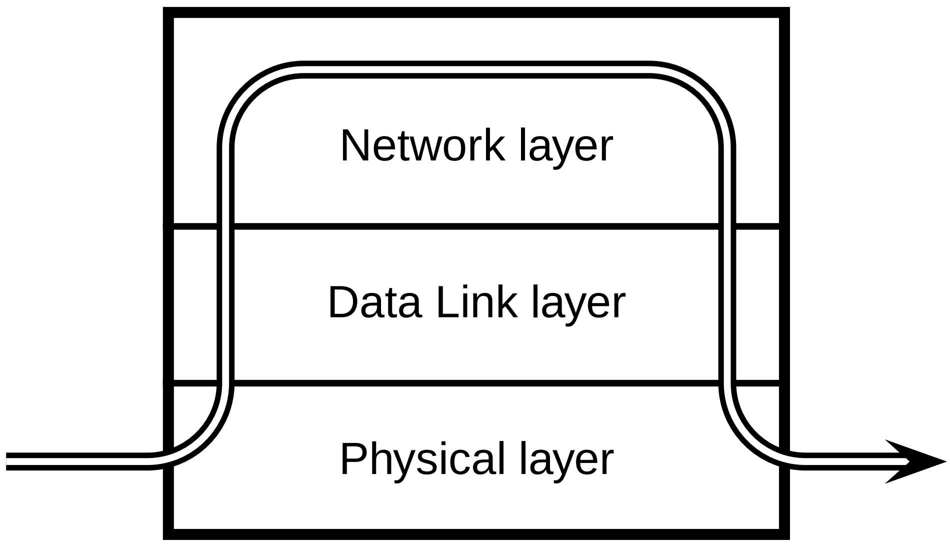
The main responsibility of the DLL protocol is to ensure an **error–free** transmission, so that higher–layer protocols can assume they are handling packets of meaningful data. This responsibility is unrelated to **reliable transport** layer, which takes care of lost packets, not corrupted bits.

Link–layer switch stack



We will need to refer to the NL stack from time to time.

Network router stack



Review of 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 the exact value for vacuum is 299,792,458 m/s).

Electrical signal propagates at the speed of 2×10^8 m/s in copper coax cable
(a bit faster in a 300Ω lead conductor).

The essential components

Data communication requires the following four components:

Message: is the collection of data items to be communicated.

Sender: is the set of devices from which the message originates.

Receiver: is the set of devices to which the message is communicated.

Transmission medium: represents the transmission technology used. In traditional systems, it can be some form of energy[a] moving through a wire, glass, air or whatever.

Senders and receivers are similar and can be lumped into a single category. We will examine the 3 different categories one at a time.

[a] This is not a Physics course and there is no room in it for theoretical Physics here.

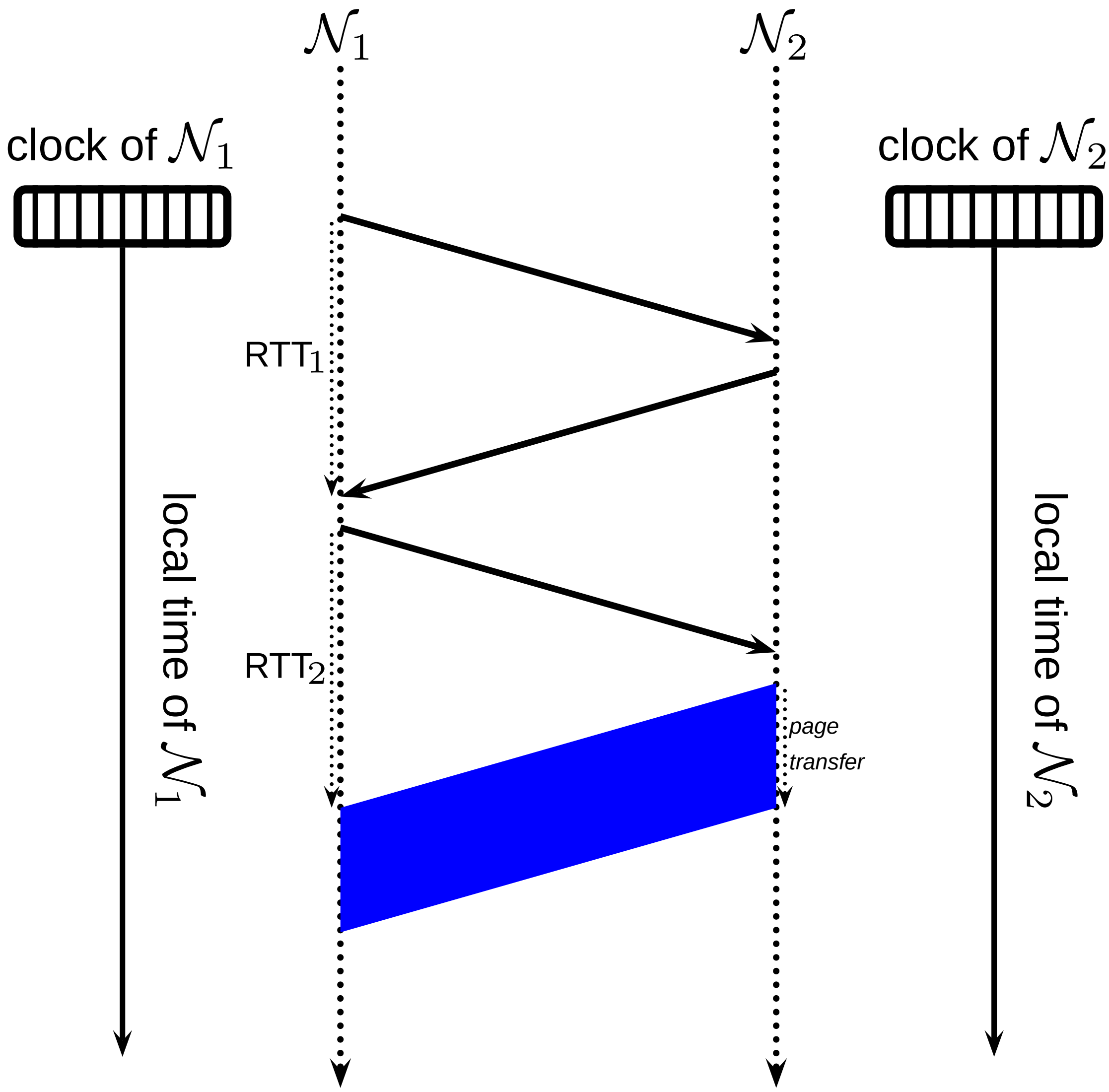
Message

This is what is being communicated. The DLL and Pl have no understanding of its meaning and concern themselves solely with a reliable delivery of the symbols forming messages.

Devices

A sender or a receiver is an autonomous device which we usually equate with a **NIC** (Network Interface Card) even though most of them are totally different internally (routers, switches, etc.).

The main property of a device is that it operates independently from other devices; this implies that it has its own clock (very important).



The two clocks tick at their own rates (just like all the other clocks) and

More on clocks

We can define a **Universal Time** (UT) as the time indicated by one chosen clock (say, at Greenwich). Let us suppose that the basic unit of time is a ps (picosecond = $10^{-12}s$), an arbitrary assumption; the UT clock ticks–by definition– at a rate of 1THz.

Consider now two devices $\mathcal{N}_1$ and $\mathcal{N}_2$ each with its own clock (let them be called $\mathcal{T}_1$ and $\mathcal{T}_2$). The clocks tick at their own rates, presumably very close to 1THz; these rates could be 1.001THz (Hz with respect to UT time) and 0.995THz respectively but, of course, each thinks that it is the correct one.

It is important to note that neither $\mathcal{N}_1$ nor $\mathcal{N}_2$ know about the UT time.

So,

$$1UTs = 1.001\mathcal{N}_1 s = 0.995\mathcal{N}_2 s$$

If $\mathcal{N}_1$ transmits 1000b at 10Gb/s (i.e. $100ps$ per bit), it will transmit them in

$$1000 \times \frac{1}{10^{10}} s \times \frac{1}{1.001} = 0.99900999 \times 10^{-7} = 99,900.1ps$$

of UT time and in $99,400.599ps$ of $\mathcal{N}_2$ time.

During the period of time when $\mathcal{N}_2$ receives the 1000 bits from $\mathcal{N}_1$, its clock will tick only 99400 times.

If $\mathcal{N}_2$ insists that its clock is the only correct one, it will think that it received only 994 bits (and some of these bits will be corrupted).

Protocols

Communication requires that all parties follow the same **communication protocol** which, in fact, is made of two unrelated parts:

- The data representation protocol which determines how to encode (and decode) data. Also called data encoding.
- The data transmission protocol which determines how data should be moved around. Also called data flow.

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. 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. USB 3.0, SATA or 10G Ethernet). Bits are typically grouped into units called **messages**, **packets** or **cells**.

Note that **byte** is not used as a unit in the context of the PL or DLL. Therefore, attention to details is required when one talks about different network layers simultaneously, e.g. a 64B message (here, Byte is the word) becomes a sequence of 976b (bit is the word) at the PL.[a]

[a] 944b = 64B + 20B (TCP) + 20B (IP) + 14B+4B (MAC).

Transmission medium

Many different ways of communicating data to the receiver can be devised. Unless performance issues are of interest, it is convenient to abstract from the exact nature of the communication process and refer to the medium connecting the sender to the receiver by a generic term: the **link**.

Thus, a link is a connector linking two (or more) devices. When two devices are involved, the link is a point–to–point link; otherwise, it is a broadcast link.

Data encoding

A lot of time will be devoted to the various ways of encoding the basic forms of digital data which are:

- Text.
- (Binary) numbers.
- Still images.
- Audio.
- Video.

Transmission medium

Many different ways of communicating data to the receiver can be devised. Unless performance issues are of interest, it is convenient to abstract from the exact nature of the communication process and refer to the medium connecting the sender to the receiver by a generic term: the **link**.

Thus, a link is a connector linking two (or more) devices. When two devices are involved, the link is a point–to–point link; otherwise, it is a broadcast link.

Data encoding

A lot of time will be devoted to the various ways of encoding the basic forms of digital data which are:

- Text.
- (Binary) numbers.
- Still images.
- Audio.
- Video.

Data flow

The movement of data from the sender to the receiver is ruled by a set of protocols. In this course we deal primarily with transmission over a single link hence the high–level protocols such as **IP**, **TCP** will not be discussed at all.

Several protocols are used; their main focus is on efficiency on the one hand and reliability on the other.

It is critical to realise that “reliable” means passing a statistical test of some sort and does not imply “correct” or “intact” or any similar thing.