

01 - Introduction

- My name is Davide Eynard
 - ✉ davide.eynard@usi.ch
- This course deals with
 - Internet
 - ... and **not only the Web!**
 - Technology
 - ... and how technology + internet are changing our lives

Università della Svizzera italiana	Facoltà di scienze della comunicazione	I
		3

■ Tentative program:

- Internet Basics (2 lessons)
- Web 1 (3 lessons)
- Web 2 (3 lessons)
- Web 3 (3 lessons)
- Monographic lessons on copyright, security, and searching

■ Student Evaluation

- Participation in classes
- Homeworks
- Project: write a monography about a topic related to the course

What is the Internet?

“The Internet is a global data communications system. It is a hardware and software infrastructure that provides connectivity between computers.”

... and it is also the place from where I took this definition :-)

Of course it is also much more, such as:

- the **physical infrastructures** used to transfer data
- the **protocols** used to manage such infrastructures
- the **services** provided by software applications that exploit it for data transfer
- the **organisations** that allow it to function properly
- the **people** that provide contents, data, and hardware

- (see <http://www.zakon.org/robert/internet/timeline>)
- 1961
 - First paper on packet-switching theory
- 1969
 - ARPANET commissioned by DoD for research into networking
 - 4 nodes: UCLA, Stanford, UCSB, University of Utah
 - First "Request for Comment" (RFC)
- 1971
 - Ray Tomlinson of BBN invents email program to send messages across a distributed network
- 1978
 - Definition of the TCP/IP protocol. This will become *the* standard for the Internet in 1983 (Jan, 1st).

- 1984
 - Number of hosts breaks 1,000
 - Domain Name System (DNS) introduced
- 1987
 - 1000th RFC
 - Number of hosts breaks 10,000
- 1990
 - Electronic Frontier Foundation is founded by Mitch Kapor
 - RFC 1149: A Standard for the Transmission of IP Datagrams on Avian Carriers
 - The World, the first dial-up provider of Internet Access, comes online
- 1991
 - World-Wide Web (WWW) released by CERN (see <http://info.cern.ch/>)

Internet Timeline

- 1992
 - Number of hosts breaks 1,000,000
- 1995
 - Traditional online dial-up systems (CompuServe, America Online, Prodigy) begin to provide Internet access
 - amazon.com, yahoo.com, ebay.com, and msn.com are launched
- 1998
 - google.com launched
- 2000
 - wikipedia.org launched
- 2005
 - First video uploaded on youtube.com

Ok, but... how does it work?

- As far as now we have dealt with the Internet... what about the technologies?
- Let's start from some examples. We will then introduce the technologies that are needed to run them and explain what happens at different levels of abstraction
 - You are chatting with a friend on MSN
 - You are downloading email from Outlook (or Thunderbird)
 - You cannot connect to a Web page and want to check what is wrong

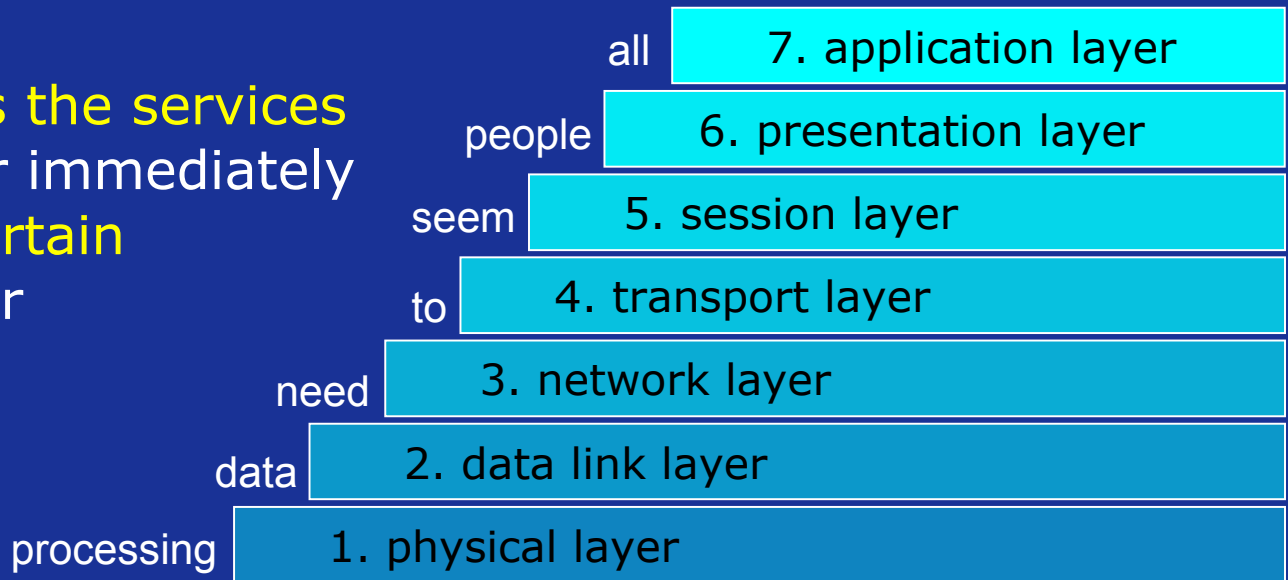
The ISO/OSI Stack

- For a network to operate correctly, it is essential that all providers of physical devices and services follow the same protocols

This can be achieved by adopting **shared standards**

- The standard named **ISO/OSI** (International Organization for Standardization/Open Systems Interconnection) is structured in **seven layers**, globally called the **ISO/OSI stack**

Every layer **exploits the services offered** by the layer immediately below and **offers certain services** to the layer immediately above



Università della Svizzera italiana	Facoltà di scienze della comunicazione	I
		10

Layers 1 and 2

- Layer 1 (the **Physical layer**) is responsible of the management of **physical devices** (cables etc.) used for transmitting data
- Layer 2 (the **Data Link layer**) exploits the services provided by the physical layer and is in charge of transmitting data from a node of a network to another node which is **directly connected** to the former node



- Every device is identified by a unique hardware MAC (Media Access Control) address, also called Ethernet Address.
- The services offered by layer 2 include the automatic detection of possible transmission errors.

Layer 3: the Network Layer

- Layer 3 (the **Network layer**) is in charge of transmitting a single packet from a source host to a destination host, which in general are not directly connected

Layer 3 is therefore responsible of the **addressing** and the **routing** of packets

- Protocols typical of Layer 3 are **IP** (**I**nternet **P**rotocol) and **ICMP** (**I**nternet **C**ontrol **M**essage **P**rotocol)

Every **network device** is given an *address* (called **IP address**) which identifies it on the network. Note: the IP address is different from the MAC address! (see following slides)

Layer 4: the Transport Layer

- Layer 4 (the **Transport layer**) is in charge of realizing a stable connection between two applications, typically running on different hosts
- As far as the internet is concerned, the most widely used layer 4 protocol is **TCP** (Transmission Control Protocol), typically coupled with **IP** at layer 3

The combined protocol, spanning layers 3 and 4, is known as **TCP/IP**

- The service provided by TCP is a **reliable flow of binary data between applications**

In particular the TCP protocol guarantees that the packets transmitted via IP **are delivered in the correct order**, and that **lost packets are sent again**

- Moving up in the ISO/OSI stack:
 - Layer 5 (the **Session layer**) is in charge of services like authentication, access control, and management of work sessions
 - Layer 6 (the **Presentation layer**) provides for services like the encryption and decryption of messages, data compression, and so on (i.e. TLS, SSL, MIME)

Layer 7: the Application Layer

- Layer 7 (the **Application layer**) includes protocols that define the interfaces between the network and the software applications running on hosts
- Examples of layer 7 protocols are:
 - **FTP** (**F**ile **T**ransfer **P**rotocol): for transferring files
 - **HTTP** (**H**ypertext **T**ransfer **P**rotocol): for transferring Web pages
 - **SMTP** (**S**imple **M**ail **T**ransfer **P**rotocol): for sending email messages
 - **POP** (**P**ost **O**ffice **P**rotocol): for the local management of email messages on a client host
 - **IMAP** (**I**nternet **M**essage **A**ccess **P**rotocol): for the management of email messages on a remote mail server
 - ...

OSI vs Internet Protocol Suite

- OSI is not the only standard to define different layers of communication between machines
- The Internet Protocol Suite is a more compact alternative to describe the organization in layers of the different protocols

OSI model

7. Application Layer

NNTP • SIP • SSI • DNS • FTP • Gopher • HTTP • NFS • NTP • SMPP • SMTP • DHCP • SNMP • Telnet • (more)

6. Presentation Layer

MIME • XDR • TLS • SSL

5. Session Layer

Named Pipes • NetBIOS • SAP • SIP • L2TP • PPTP

4. Transport Layer

TCP • UDP • SCTP • DCCP

3. Network Layer

IP • ICMP • IPsec • IGMP • IPX • AppleTalk

2. Data Link Layer

ARP • CSLIP • SLIP • Ethernet • Frame relay • ITU-T G.hn DLL • PPP

1. Physical Layer

RS-232 • RS-449 • V.35 • V.34 • I.430 • I.431 • T1 • E1 • POTS • SONET/SDH • OTN • DSL • 802.11a/b/g/n PHY • 802.15.x PHY • ITU-T G.hn PHY • Ethernet • USB • Bluetooth

This box: [view](#) • [talk](#) • [edit](#)

Internet Protocol Suite

Application Layer

BGP • DHCP • DNS • FTP • HTTP • IMAP • IRC • LDAP • MGCP • NNTP • NTP • POP • RIP • RPC • RTP • SIP • SMTP • SNMP • SSH • Telnet • TLS/SSL • XMPP • (more)

Transport Layer

TCP • UDP • DCCP • SCTP • RSVP • ECN • (more)

Internet Layer

IP (IPv4, IPv6) • ICMP • ICMPv6 • IGMP • IPsec • (more)

Link Layer

ARP/InARP • NDP • OSPF • Tunnels (L2TP) • PPP • Media Access Control (Ethernet, DSL, ISDN, FDDI) • (more)

This box: [view](#) • [talk](#) • [edit](#)

- You are chatting with a friend on MSN or some other IM system
 - At the application layer, you are using a protocol like XMPP (or a proprietary one)
 - Few systems also provide a secure connection at the presentation layer with SSL
 - At transport layer, UDP is mostly used
 - At the network layer, IP is used: your messages leave from a machine identified by a specific IP address and they are addressed to another IP
 - Below this, at the data link layer, an ethernet connection has been established in your home (between your computer and the ADSL router), and a PPP connection has been established between the ADSL router and the Internet Service Provider (ISP)
 - At the physical layer, you have connected your computer to the router with a specific cable

- You are downloading email from a mail client application
- Let us suppose you are connecting with USI Wireless and go from the bottom up:
 - The physical layer changes, as you are connecting with WiFi instead of a cable
 - Same holds for the data link layer: we can suppose that USI servers are directly connected to the Internet through a LAN
 - The network layer is the same: the protocols we need for email exchange rely on IP
 - At the transport layer, TCP is used as we need a more complex type of connection
 - At the presentation layer, we can rely on TLS/SSL for account authentication and secure email transfer
 - At the application layer, we use POP3/IMAP to download our email messages

- You cannot load a Web page and want to check what is wrong...
- Start from the bottom up!
 - Physical: check the cable/wifi connection
 - Data link: check if the ADSL router is working
 - Network: check that you got a valid IP address
 - Transport: check that the server you are connecting to is reachable
 - Presentation: check that you have the correct certificates to perform a secure connection (if needed)
 - Application: check that your browser is not in Offline mode :-)

- An IP address is comprised of **four bytes**, typically represented by four natural numbers in decimal notation (each number between **0** and **255**) separated by dots, like for example

195.176.39.49

NOTE: the fact that this is the *typical* form of an IP address does not mean it is the *only* one... Just keep in mind that an IP is a “big”, 32 bits (4 bytes) number, that can be encoded in different ways

- The special IP address **127.0.0.1** is conventionally used in every host as the IP address of the host itself. The symbolic IP address **localhost** is equivalent to **127.0.0.1** (who does this conversion?)

- An IP address may be assigned to a network device **statically** or **dynamically**
- A **static IP address** is assigned to a network device once and for all (and thus behaves as a telephone number, which is assigned to a specific telephone line, or to a specific SIM card in the case of mobile phones)
- On the contrary, a **dynamic IP address** is temporarily assigned to a network card for the duration of a working session

For example, when one connects to the internet with a PC through a commercial provider, the provider temporarily assigns to the PC's network card an IP address, which is then used for the whole working session

Domain Name Server

- Names are much easier to remember than IP addresses
 - google.com or 74.125.43.*whatever*?
- Historically, domain name resolution has its origins in a textual file called "hosts"
 - That file is still present on your computers... find it and play with it!
- RFC 625 (March, 1974) settled that the Stanford Research Institute Network Information Center (NIC) would serve as the official source of the master hosts file. DNS as we know it, however, entered the scene only 10 years later
 - What does it mean to have a centralized hosts file?
 - Actually, the current DNS system is more distributed...

Domain Name Server (2)

- Every time you connect to a website by name, you are transparently connecting to a DNS server which resolves that name to the server's IP
 - One more reason why you cannot connect to a website (see Example 3): DNS is not working/accessible
- A domain name consists of one or more parts, called *labels*, delimited by dots (i.e. `www.google.com`)
 - The rightmost label is the *top-level domain*
 - Each label to the left represents a *subdomain* of the domain to the right
 - A *hostname* is a *domain name* that has at least one IP address associated (i.e. "com" is not a hostname, while `google.com` and `www.google.com` are)
- Why are some website names *not* beginning with "www"?
 - Because it is *not compulsory* and it is up to the website maintainer's choice!

- The reliable flow of data realized by TCP/IP takes place between certain “virtual locations”, called **sockets**, which are used by software applications to exchange data
- A socket is comprised of:
 - the **IP address** of a network card
 - a **TCP port**, which may be regarded as a virtual location at which data are delivered (note that a port is a software entity, not a physical device!)

For example, socket

195.176.39.49:80

combines the IP address **195.176.39.49** (which identifies a network card) and the TCP port **80**, typically used by the **http protocol** to deliver the requests of transmission of Web pages

■ Some Web references:

■ Internet timelines:

<http://www.zakon.org/robert/internet/timeline/>

http://www.xiongdudu.com/image/Internet_Timeline

<http://www.niallkennedy.com/blog/timelines/youtube/>

■ Internetworking layers:

http://en.wikipedia.org/wiki/OSI_model

http://en.wikipedia.org/wiki/Internet_Protocol_Suite

<http://www.wilsonmar.com/1isotp.htm>

■ The Web is dead:

http://www.wired.com/magazine/2010/08/ff_webrip/all/1

<http://www.boingboing.net/2010/08/17/is-the-web-really-de.html>

■ Zen and the Art of the Internet (a beginner's guide to the Internet, first edition, January 1992)

http://www.cs.indiana.edu/docproject/zen/zen-1.0_toc.html

■ Some tools:

■ The “ping” command (i.e. “ping www.google.com”)

■ The “whois” command (i.e. “whois eynard.it”, “whois 88.149.156.170”)