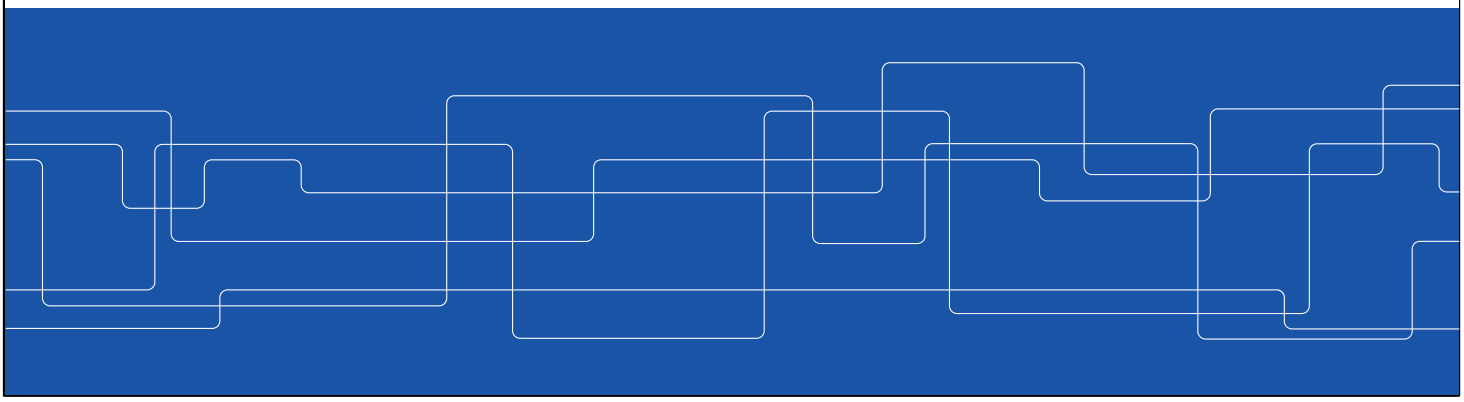
KTH ROYAL INSTITUTE
OF TECHNOLOGY

Networks and Interprocess Communication

Vladimir Vlassov and Johan Montelius



Requirements

- Performance
- Scalability
- Reliability
- Security
- Mobility
- Quality of Service
- Multicasting

The main non-functional properties of systems that affect the **quality of the service** experienced by clients and users are **reliability, security, and performance**. The performance aspect of quality of service was initially defined in terms of responsiveness and computational throughput, but it has been **redefined regarding the ability to meet timeliness guarantees**. **Quality of service includes the ability to meet deadlines when transmitting and processing streams of real-time multimedia data.** Its achievement depends upon the availability of the necessary computing and network resources at the appropriate times. This imposes significant new requirements on computer networks. Applications that transmit multimedia data **require guaranteed bandwidth and bounded latencies for the communication channels** that they use. Some applications vary their demands dynamically and specify a minimum acceptable quality of service and the desired optimum.

Network. Internet

A ***network*** is a hardware and software data communication system that provides interconnection of computers and other devices.

Internet is a set of networks connected with routers.

The Internet is the largest internet that includes commercial, military, university, and other networks with different physical links and various protocols, including IP (Internet Protocol)

Types of networks

- WAN - Wide Area Networks
- MAN - Metropolitan Area Networks
- LAN - Local Area Networks
- PAN - Personal Area Networks

Latency

Transfer rate:

What is the rate at which we can send data?

Performance

- Latency - how long time does it take to send an empty message?
- Transfer rate - what is the rate at which we can send data?

Latency

Why does it take time to send a message?

- distance - speed of signal (light)
- access - granting of resource
- routing - processing in nodes

fast as ..

What is the speed of light?
300 000 km/s ... or 300 km/ms

Distance in ms:
Stockholm - Hamburg approx. 800 km or 3 ms
Stockholm - NYC approx. 6.600 km or 23 ms
Stockholm - Melbourne approx. 15.600 km or 52 ms

Routers. switches and fiber optics adds to this so Melbourne is approx. 300 ms away.

ping

```
vladv — bash — 80×24
pc65:~ vladv$ ping www.aflcommunityclub.com.au
PING www.aflcommunityclub.com.au (202.74.66.109): 56 data bytes
64 bytes from 202.74.66.109: icmp_seq=0 ttl=43 time=371.140 ms
Request timeout for icmp_seq 1
64 bytes from 202.74.66.109: icmp_seq=2 ttl=43 time=406.258 ms
64 bytes from 202.74.66.109: icmp_seq=3 ttl=43 time=626.502 ms
64 bytes from 202.74.66.109: icmp_seq=4 ttl=43 time=543.209 ms
64 bytes from 202.74.66.109: icmp_seq=5 ttl=43 time=461.641 ms
64 bytes from 202.74.66.109: icmp_seq=6 ttl=43 time=382.349 ms
64 bytes from 202.74.66.109: icmp_seq=7 ttl=43 time=611.176 ms
64 bytes from 202.74.66.109: icmp_seq=8 ttl=43 time=367.338 ms
64 bytes from 202.74.66.109: icmp_seq=9 ttl=43 time=367.141 ms
64 bytes from 202.74.66.109: icmp_seq=10 ttl=43 time=683.341 ms
64 bytes from 202.74.66.109: icmp_seq=11 ttl=43 time=605.175 ms
64 bytes from 202.74.66.109: icmp_seq=12 ttl=43 time=520.319 ms
^C
--- www.aflcommunityclub.com.au ping statistics ---
13 packets transmitted, 12 packets received, 7.7% packet loss
round-trip min/avg/max/stddev = 367.141/495.466/683.341/112.186 ms
pc65:~ vladv$
```

Using ICMP packages might give a better value, UDP might be slower.

Ping is a computer network administration software utility used to test the reachability of a host on an Internet Protocol (IP) network. Ping measures the round-trip time for messages sent from the originating host to a destination computer that are echoed back to the source. The name comes from active sonar terminology that sends a pulse of sound and listens for the echo to detect objects underwater. Ping operates by sending Internet Control Message Protocol (ICMP) echo request packets to the target host and waiting for an ICMP echo reply.

The Internet Control Message Protocol (**ICMP**) is a supporting protocol in the Internet protocol suite. It is used by network devices, including routers, to send error messages and operational information indicating, for example, that a requested service is not available or that a host or router could not be reached.

Latency in different networks

- LAN/WLAN - local area networks (Ethernet/WiFi) 1 - 10 ms
- WAN - wide area networks (IP routed) 20 - 400 ms
- Mobile networks 40 - 800 ms
  5G: 20-30 ms
- Satellite (geo-stationary) > 250 ms

Message size

How does latency vary with the size of the messages?

- The packet delivery time or latency is the time from when the first bit leaves the transmitter until the last is received.
- In the case of a physical link, it can be expressed as:
  Packet delivery time = Transmission time + Propagation delay
  - where
    - Transmission time = Packet size / Bit rate
      - The transmission time should not be confused with the propagation delay, which is the time it takes for the first bit to travel from the sender to the receiver.
    - Propagation time = Distance / propagation speed

The packet delivery time or latency is the time from when the first bit leaves the transmitter until the last is received. In the case of a physical link, it can be expressed as:
Packet delivery time = Transmission time + Propagation delay
Where Packet transmission time = Packet size / Bit rate
The transmission time should not be confused with the propagation delay, which is the time it takes for the first bit to travel from the sender to the receiver.
Propagation time = Distance / propagation speed

Transfer rate

The rate at which we can send data (does not mean that it has arrived).

What is the transfer rate of:

ADSL	1 - 20 Mb/s
Ethernet	100 Mb/s - 1 Gb/s
802.11	11 Mb/s, 54 Mb/s, 72 Mb/s ...
3G/4G	1 Mb/s, 2 Mb/s, ... 100 Mb/s
5G	over 1,000 Mb/s (1Gb/s).

Is this shared with others?

Older 2G connections give a download speed of around 0.1Mbit/s, rising to around 8Mbit/s on the most advanced 3G networks. Rates of around 60Mbit/s are available on 4G mobile networks in the UK (but this can be substantially higher in other countries like the US). Next-generation 5G mobile networks target a download speed of over 1,000Mbit/s (1Gbit/s).

Overhead

medium access: 802.11 – RTS/CTS
error handling: detection, forward error correction, ARQ
header: MAC header, IP header, TCP ...
flow control: TCP window

IEEE 802.11 is a set of media access control (MAC) and physical layer (PHY) specifications for implementing wireless local area network (WLAN) computer communication in the 900 MHz and 2.4, 3.6, 5, and 60 GHz frequency bands.
RTS/CTS (Request to Send / Clear to Send) is the **optional mechanism used by the 802.11** wireless networking protocol **to reduce frame collisions** introduced by the hidden node problem.
Automatic Repeat request (**ARQ**), also known as **Automatic Repeat Query**, is an **error-control method** for data transmission that uses acknowledgments (messages sent by the receiver indicating that it has correctly received a data frame or packet) and timeouts (specified periods allowed to elapse before an acknowledgment is to be received) to achieve reliable data transmission over an unreliable service. Suppose the sender does not receive an acknowledgment before the timeout. In that case, it usually re-transmits the frame/packet until the sender receives an acknowledgment or exceeds a predefined number of re-transmissions.

In computer networking, **RWIN (TCP Receive Window)** is the amount of data a computer can accept without acknowledging the sender. If the sender has not received acknowledgment for the first packet it sent, it will stop and wait, and if this wait exceeds a certain limit, it may even retransmit. This is how TCP achieves reliable data transmission.

What's in it for me?

The application layer transfer rate is much lower than the physical layer bit rate.

How does the application layer latency differ from the network layer latency?

Latency and transfer rate

Stockholm to Gothenburg - 400 km, best possible data communication layer?

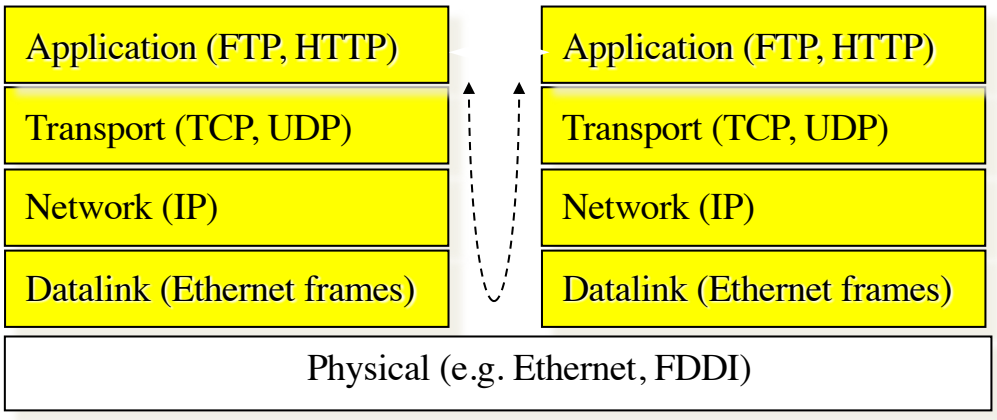
100 m^3 or five million BlueRay 50Gbyte disks, delivered in 6 h, two trucks every day

10 Gbit/s

2006 – Nippon Telegraph and Telephone Corporation transferred 14 terabits per second over a single 160 km long optical fiber
2009 – Bell Labs in Villarceaux, France transferred 100 Gbit/s over 7000 km fiber
2010 – Nippon Telegraph and Telephone Corporation transferred 69.1 Tbit/s over a single 240 km fiber multiplexing 432 channels, equating to 171 Gbit/s per channel
2012 – Nippon Telegraph and Telephone Corporation transferred 1 Petabit per second over 50 kilometers over a single fiber

Multi-Layered Network Architecture

The seven-layer OSI (Open System Interconnect) model
The IP networking stack includes 5 layers



Fiber Distributed Data Interface (FDDI) is a standard for data transmission in a local area network. It uses optical fiber as its standard underlying physical medium. However, it was also later specified to use copper cable, in which case it may be called CDDI (Copper Distributed Data Interface), standardized as TP-PMD (Twisted-Pair Physical Medium-Dependent), also referred to as TP-DDI (Twisted-Pair Distributed Data Interface). FDDI was effectively made obsolete in local networks by Fast Ethernet, which offered the same 100 Mbit/s speeds but at a much lower cost, and, since 1998, by Gigabit Ethernet due to its speed, even lower price, and ubiquity

Communication layers

Application	the end product
Presentation	encoding of information, serialization, marshaling
Session	security, authentication, initialization
Transport	messages, streams, reliability, flow control
Network	addressing of nodes in a network, routing, switching
Data link	point to point deliver of frames, medium access, link control
Physical layer	bits to analog signals, electrical, optical, radio ...

Internet stack

HTTP, FTP, SMTP

TCP, UDP, SCTP, ICMP

IP, ARP

Ethernet, WiFi, ..

ICMP: Internet Control Message Protocol
SCTP: Stream Control Transmission Protocol

What if

What would the world look like ...

.. if we only had Ethernet?

Range: 1-2 km
Bandwidth: 10-10,000 Mbps
Latency 1-10 ms

Routing

Two approaches:

- Distance vector: send routing table to neighbors, RIP, BGP
- Link state: tell everyone about your direct links, OSPF

Pros and cons?

A routing algorithm has two parts (1) It must make decisions that determine the route taken by each packet as it travels through the network; (2) It must dynamically update its knowledge of the network based on traffic monitoring and the detection of configuration changes or failures.
Vector distance – the number of hops to the given destination. A broken link has the value “infinity.”
Distance-vector – slow convergence;
The Routing Information Protocol (RIP) is one of the oldest distance-vector routing protocols, which employs the hop count as a routing metric. Send a routing table to neighbors each 30 sec. Broken
Border Gateway Protocol (BGP) is a standardized exterior gateway protocol designed to exchange routing and reachability information among autonomous systems (AS) on the Internet.
OSPF: Open Shortest Path First – converges more rapidly than RIP

IP addresses

What is the structure of an IP address?

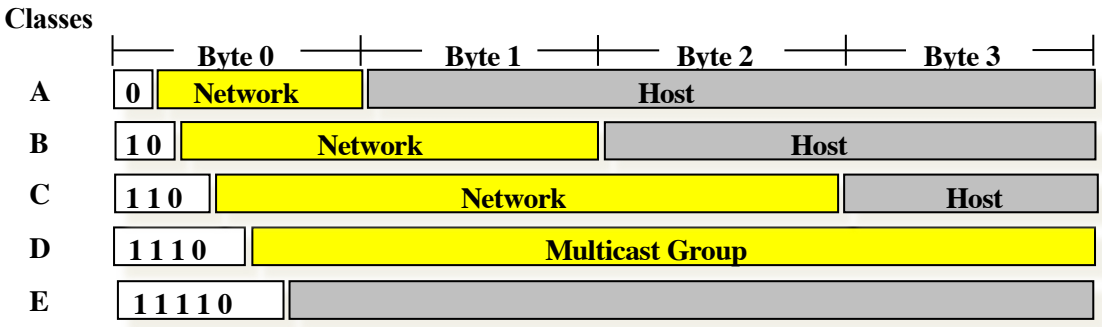
How would you allocate IP addresses to make routing easier?

What is happening?

Every IP address—such as 76.240.249.145—is divided into two sections that define 1) your network and 2) your computer or host. Those two sections comprise the basic structure of IP addresses: the network ID and the host ID. All computers on the same network share the same network ID. Classes A, B, C, D, E.

IP Address Classes (Classful addressing)

A (1-126.x.x.x) – 126 address blocks, each of 16,000,000 addresses.
B (128-191.x.x.x) – one address block contains ~65,000 addresses.
C (192-223.x.x.x) – one address block contains 254 addresses.
D (224-239.x.x.x) – multicast addresses.
E (240-255.x.x.x) –reserved.



Classful Routing and Classless Routing

Recently, the classful addressing with five classes A-E in IPv4 has become obsolete and replaced with classless addressing.

- to tackle the problem of waste/lack of IP addresses

Classful routing: an address is divided into three parts: Network, Subnet, and Host

Classless routing: an address is divided into two parts: Subnet and Host

UDP and TCP

One word that describes the difference between UDP and TCP.

error-checking??

UDP and TCP

Introduces two communication abstractions:

- UDP: datagram
- TCP: stream

- Gives us port numbers to address processes on a node.
- About hundred other protocols defined using IP. (ICMP, IGMP, RSVP, SCTP...)
- More protocols defined on top of UDP and TCP.

The Internet Control Message Protocol (**ICMP**) is a supporting protocol in the Internet protocol suite. It is used by network devices, including routers, to send error messages and operational information indicating, for example, that a requested service is not available or that a host or router could not be reached.
The Internet Group Management Protocol (**IGMP**) is a communications protocol used by hosts and adjacent routers on IPv4 networks to establish multicast group memberships. **IGMP** is an integral part of IP multicast.
The Resource Reservation **Protocol** (**RSVP**) is a transport layer **protocol** designed to reserve resources across a network for an integrated services Internet.
In computer networking, the Stream Control Transmission Protocol (**SCTP**) is a transport-layer protocol, serving in a similar role to the popular protocols TCP and UDP. It is standardized by IETF in RFC 4960

UDP

- A datagram abstraction, independent messages, limited in size.
- Low cost, no set up or tear down phase.
- No acknowledgment.

TCP

- A duplex stream abstraction.
- Reliability, lost or erroneous packets are retransmitted.
- Flow control to prevent the sender from flooding the receiver.
- Congestion friendly, slows down if a router is choked.

UDP and TCP

- UDP: small size messages, build your streams
- TCP: large size messages, flow control of a stream of messages

Can you trust TCP delivery?

Sockets

A socket is the programmer's abstraction of the network layer
- an endpoint of a virtual network connection;
- identified by an IP address & port number, and a transport protocol (TCP, UDP, …)
  - Datagram sockets for messages (UDP)
  - Stream sockets for duplex byte streams (TCP)

Sockets, a.k.a. Berkeley sockets, were introduced in 1981 as the Unix BSD 4.2 generic API for inter-process communication.
- Earlier, a part of the kernel (BSD Unix)
- Now, a library (Solaris, MS-DOS, Windows, OS/2, MacOS)

Stream Socket

A TCP socket for stream-based communication

- Server
  - Creates a listening socket bound to a port (could be in several steps: create, bind, listen)
  - Accepts an incoming connection request and creates a communication socket for reading/writing a byte stream.
- Client
  - Creates a communication socket and connects it to a server identified by an IP address and a port.
  - Reads/writes from a socket.

A Server in Erlang

```
init(Port) ->
   case gen_tcp:listen(Port, [..]) of
      {ok, Listen} ->
         handler(Listen),
         gen_tcp:close(Listen);
      {error, Error} ->
         error
   end.
```

```
handler(Listen) ->
   case gen_tcp:accept(Listen) of
      {ok, Client} ->
         request(Client),
         handler(Listen);
      {error, Error} ->
         error
   end.
```

A Server in Erlang

```
request(Client) ->
   case gen_tcp:recv(Client, 0) of
      {ok, Request} ->
         Response = reply(Request),
         gen_tcp:send(Client, Response);
      {error, Error} ->
         error
      end,
   gen_tcp:close(Client).
```

```
reply(Request) ->
   :
   generate and return
   a byte sequence
```

Datagram socket

- Server
  - Create a message socket and bind it to a port.
  - Receive an incoming message (message contains a source IP address and port number).

- Client
  - Create a message socket bound to a source port.
  - Create a message and give it a destination address and port number.
  - Send the message.

Marshaling of data

How do we transform internal data structure into the sequencing of bytes?

- Language dependent: Java serialization, Erlang external term format
- Independent: XML, Google Protocol Buffer, ASN.1
  - message format defined by specification: XML Schema, .proto, ...
  - A compiler uses the specification to generate an encoder and decoder

.proto - Google developed Protocol Buffers for use internally and has provided a code generator for multiple languages under an open-source license. A software developer defines data structures (called messages) and services in a proto definition file (.proto) and compiles it with protoc.

Example

ASN.1 specification

```
FooProtocol DEFINITIONS ::= BEGIN
   FooQuestion ::= SEQUENCE {
      trackingNumber INTEGER,
      question IA5String}
   FooAnswer ::= SEQUENCE {
      questionNumber INTEGER,
      answer BOOLEAN}
END
```

C data structures

```
struct foo_question {
         int tracking_number;
         char question[128];
}

foo = {5, "Anybody there?"};
```


ASN.1 is a formal notation used for describing data transmitted by telecommunications protocols, regardless of language implementation and physical representation of these data, whether complex or very simple, whatever the application.

Summary

In a perfect world, the application layer should be independent of underlying layers.

The world is not perfect.

Understanding underlying network characteristics is essential when developing distributed applications.

ID2201 Distributed Systems

The lecture continues 16:15