MEASURING PERFORMANCE

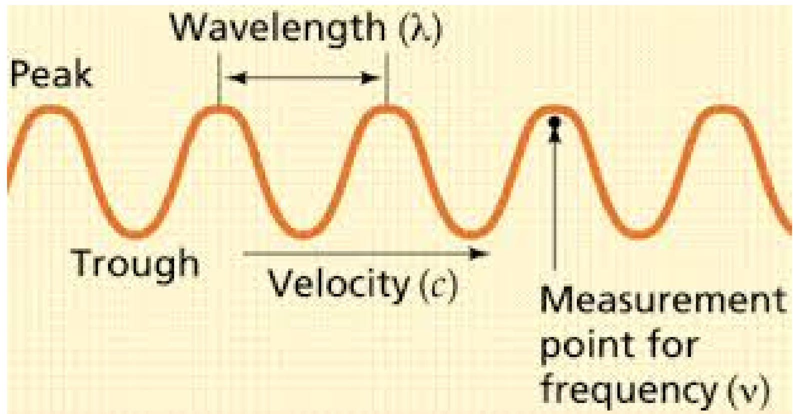
Frequency and Period

- Frequency measures how often a repeating event occurs.
- Measuring it is accomplished by counting the number of times that event occurs within a specific time period, then dividing the count by the length of the time period.
- *Frequency* f (measured in Hz) is the number of occurrences of a repeating event per unit time.
- The *period* T (measured in *sec*) is the duration of one cycle in a repeating event, so the period is the reciprocal of the frequency.
- Therefore we have the basic relation between period and frequency

$$T = \frac{1}{f}$$

Wavelength

- *Wavelength* of a sinusoidal wave is the spatial period of the wave, i.e., the distance over which the wave's shape repeats.



- From physics we know

$$\lambda = vT$$

- Thus, for periodic waves, frequency has an inverse relationship to the concept of wavelength; simply, frequency is inversely proportional to *wavelength* λ (lambda).

Frequency for Periodic Waves

- The frequency f is equal to the phase velocity v of the wave divided by the wavelength λ of the wave:

$$f = \frac{v}{\lambda}.$$

- In the special case of electromagnetic waves moving through a vacuum, then $v = c$, where c is the speed of light in a vacuum, and this expression becomes:

$$f = \frac{c}{\lambda}.$$

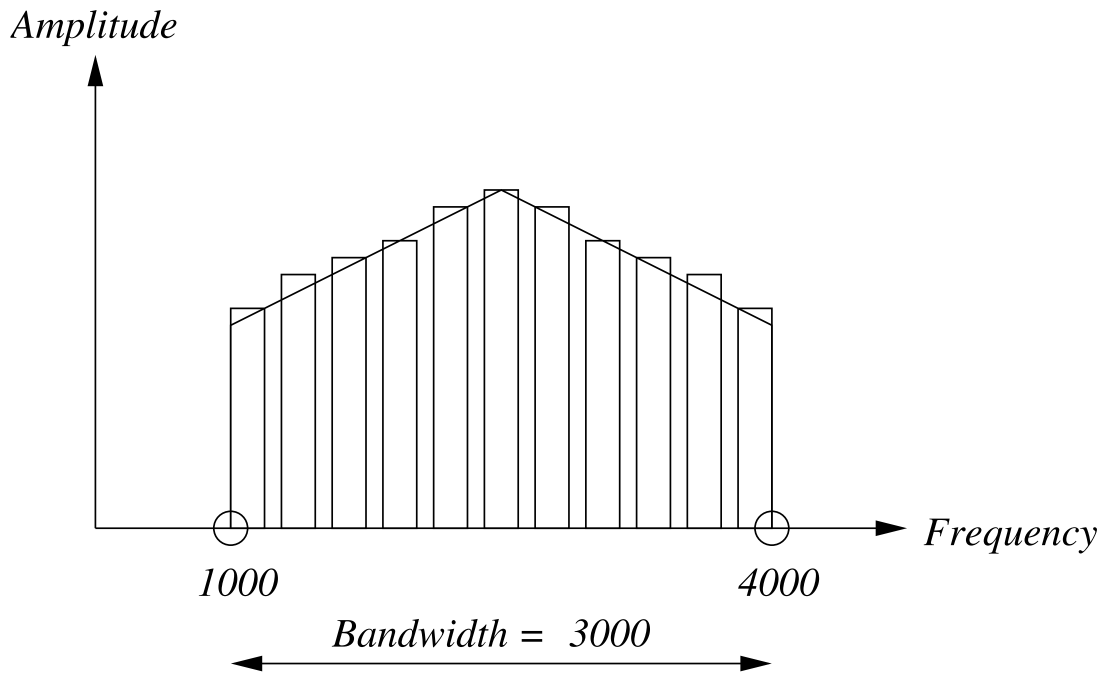
- Note the interdependence between wavelength, frequency, and speed of medium.

Network Performance

- Network performance is measured using two parameters
  - **Bandwidth**, and
  - **Latency** (also called **Delay**).
- Both delay and bandwidth depend on the medium.
- Intuitively, bandwidth measures "length" (but can be measured in Hz), while "delay" measures time, but there many types of delay.

Bandwith

- Bandwidth is a property of the medium.



- **Bandwidth** is the difference between the highest and lowest frequencies that the medium can "pass" satisfactorily.

Bandwith: Examples

- Different bands may be used per application.
  - But you must fit the application within the selected bands!
- Voice's spectrum is between 300 and $3,300$ units of frequency and so the bandwidth is $3,000$.
  - Within this band you are interested to know how many bits can be transmitted per second.
- Different applications have different bandwidth requirements.
  - Instant messaging conversation: less than 1,000 bps;
  - VoIP requires 56 Kbps to sound smooth and clear;
  - Standard definition video (480p) works at 1 Mbps.
  - HD video (720p) wants around 4 Mbps, and
  - HDX video more than 7 Mbps.

Bandwidth

- **Bandwidth** is also measured as "throughput", i.e., in "# of bits per unit time", when we don't care about the actual bands being used;
  - it is a measure of the width of the frequency band.
- **Example:** the Bandwidth of a network is
  1. 100Mbps (able to deliver 100 million bits per second)
  2. 100 MBps (able to deliver 100 million bytes per second)
- Sometimes we think of bandwidth in terms of "how long it takes to transmit one bit".
- **Example:** on 100 Mbps network it takes 0.01 μs (microseconds) to transmit a bit.

Measuring Conventions

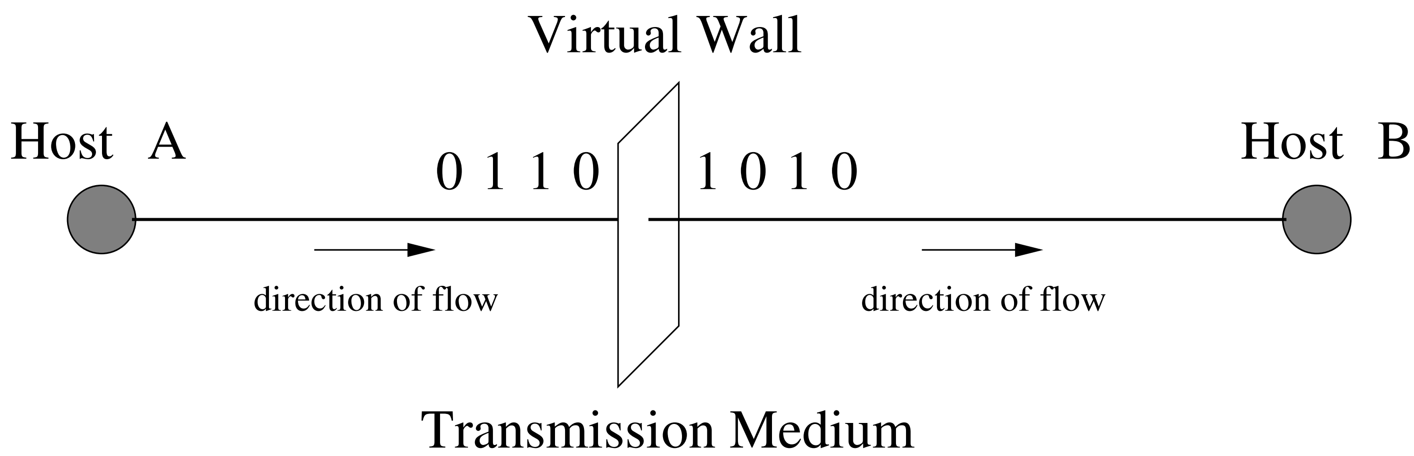
- We use
  - b for bit, and B for Byte (8 bits).
  - K for kilo, M for Mega, G for Giga.
- In measuring quantities we assume $10^3 \approx 2^{10}$.
- We use powers of 2 for bits and powers of 10 for frequencies!
  - Depending on where it is being used K can mean either 10^3 or 2^{10}, and M can mean either 10^6 or 2^{20}, etc.
- Although measured in Mbps, bandwidth is governed by the clock speed which is *pacing the transmission*.
  - A clock is measured in Hz. So in 1 Hz we can transmit 1 bit. Therefore, 10 MHz bandwidth is the same as 10 Mbps.

International System of Units (ISU)

Unit	Abbreviation	Value
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
mili	m	10^{-3}
deca	da	10^{1}
kilo	k or K	10^{3}
mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}
peta	P	10^{15}

Throughput

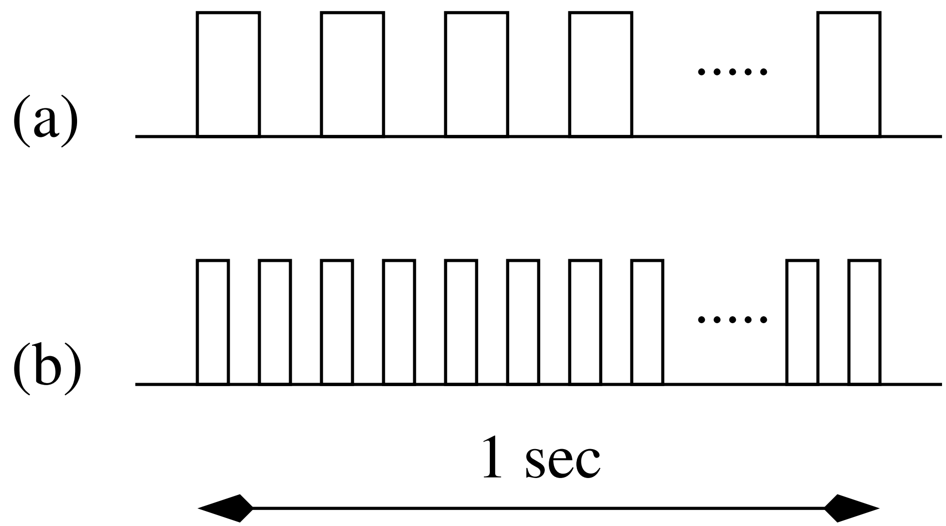
- Throughput is the measurement of how fast data can pass through a point.



- If we consider such a "measuring point" as a virtual wall then throughput is the number of bits that pass through the wall per second.
- In a way, *throughput is bandwidth that refers to measured performance.*

Distance and Time

- If you think of time as distance and the bit as a pulse of a certain width then bandwidth is how many bits fit in a unit distance.



- **Example:**
  - in picture (a) we have bits transmitted in a 1 Mbps line,
  - in picture (a) we have bits transmitted in a 2 Mbps line.

Example: Delay

- An upper bound (or best case) on latency in a medium is determined by the speed of light in that medium.
- This speed varies per medium.

Medium	Speed of Light
Vacuum	$3.0 \cdot 10^8$ m/s
Cable	$2.8 \cdot 10^8$ m/s
Fiber	$2.0 \cdot 10^8$ m/s

Common Sense Examples

- A voice grade telephone supports a frequency band ranging from 300 Hz to $3,300$ Hz. Its bandwidth will be

$$3,300 - 300 = 3,000\ Hz.$$

- When refering to communication links bandwidth refers to the number of bits per second that can be transmitted in that link.
  - The bandwidth of Ethernet is 10 $Mbps$.
  - This can vary per technology used.
- Throughput is bandwidth that refers to measured performance.
- Bandwidth requirements of an application is the number of bits per second that it needs to transmit over the network in order to have acceptable performance.

Latency and RTT

- **Latency** (or **Delay**)
  - is "how long it takes a message to travel from one end of a network to another" and is measured in time units.
- E.g., the latency of a network might be 10 ms, i.e., it takes 10 ms to travel from one end to another.
- **Round Trip Time** (abbreviated RTT)
  - measures the time for a message to reach from one end to the other and back.

Bandwidth: Example

- *How many bits can a transcontinental channel hold if it has one-way latency of* 60 *ms and bandwidth of* 50 *Mbps?*
- You find the number of bits if you multiply the latency by the bandwidth.
  - Latency: $60\ ms = 60 \times 10^{-3}\ sec$.
  - Bandwidth: $50\ Mbps = 50 \times 10^{6}\ bits\ per\ sec$.
  - Therefore we have

$$\begin{aligned}
\#\text{ of bits} &= (60 \times 10^{-3} sec) \times (50 \times 10^{6}\ bits\ per\ sec) \\
&= (60 \times 50) \times 10^{6-3}\ bits \\
&= 3 \times 10^{6}\ bits \\
&= 3\ Mb
\end{aligned}$$

Transmission Media: Parameters

- **Speed:** Max number of bits per sec that can be transmitted reliably.
- **Attenuation:** Tendency of a signal to become weak or distorted over distance.
  - Signal absorbed/dissipated during transmission.
- **Electromagnetic Interference (EMI):** Susceptibility of medium to external electromagnetic energy which is inadvertently introduced onto a link
  - causes "static audio" and "visual snow" in corresponding media.
- **Cost:** Materials plus installation.

Transmission Media: Comparison

Medium	Cost	Speed (in bps)	Attenuation	EMI	Security
UTP	Lo	1-100 M	Hi	Hi	Lo
STP	Mo	1-150 M	Hi	Mo	Lo
Coax	Mo	1 M-1 G	Mo	Mo	Lo
Optical Fiber	Hi	10 M-2 G	Lo	Lo	Hi
Radio	Mo	1-10 M	Lo-Hi	Hi	Lo
Microwave	Hi	1 M-10 G	Va	Hi	Mo
Satellite	Hi	1 M-10 G	Va	Hi	Mo
Cellular	Hi	9.6-19.2 K	Lo	Mo	Lo

Hi = High, Mo = Moderate, Va = Variable, Lo = Low

UTP = Unshielded Twisted Pair, STP = Shielded Twisted Pair

How Long Does it Take?

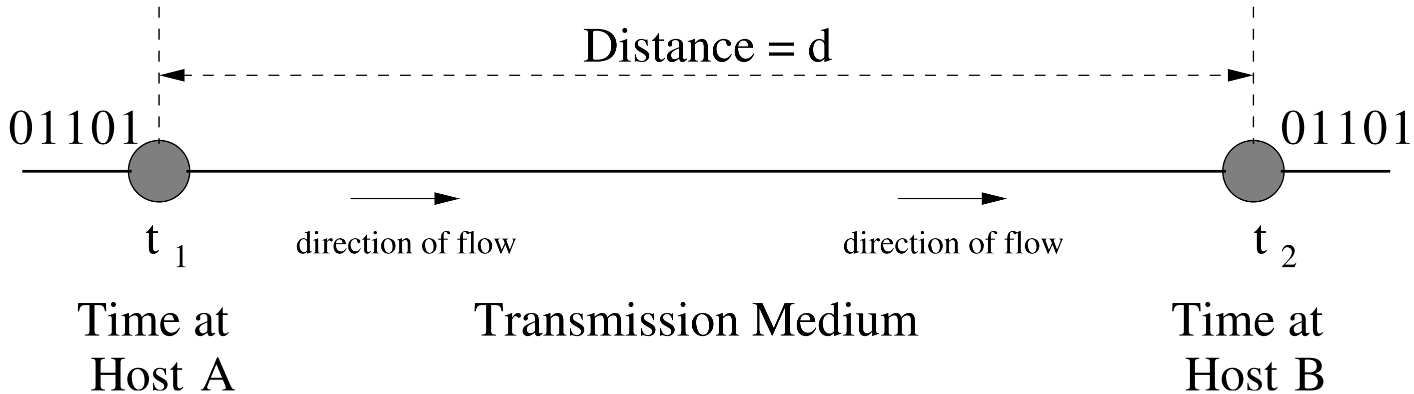
- The time it takes to transmit a unit of data in a network depends on the network bandwidth and data-unit (or packet) size.
- There are also Queuing delays in a network due to buffering, and switching.

$$
\begin{aligned}
\text{Delay} &= \text{Propagation} + \text{Transmit} + \text{Queue} \\
\text{Propagation-Delay} &= \frac{\text{Distance}}{\text{Speed of Light}} \\
\text{Transmit-Delay} &= \frac{\text{Packet Size}}{\text{Bandwidth}}
\end{aligned}
$$

- Bandwidth and Latency define and dominate the performance characteristics of a network.

Propagation Speed/Time

- **Propagation speed**
  - measures "the distance a bit can travel in one second".



- **Propagation time**
  - measures the time it takes for a bit to travel from one end to another.

- From physics we know:

$$\text{Distance} = \text{Propagation speed} \times \text{Propagation time}$$

Examples

- The propagation time (normalized in kilometers) for Twisted Pair is

$$\begin{aligned}\text{Propagation time} &= 1000\ m/(2 \times 10^8 m/s)\\ &= 5 \times 10^{-6} s\\ &= 5\ \mu s\end{aligned}$$

- The propagation time (normalized in kilometers) for Coaxial or Fiber Optic Cable is

$$\begin{aligned}\text{Propagation time} &= 1000\ m/(3 \times 10^8 m/s)\\ &= 3.33 \times 10^{-6} s\\ &= 3.33\ \mu s\end{aligned}$$

Latency and Bandwidth

- When combined they "determine" network performance.
- Consider a client and a server that exchange messages.
- Is there a (noticeable) difference between transmitting across
  - the room with 1 ms RTT, and
  - a transcontinental channel with 100 ms RTT?
- Which of the two is more important:
  - Latency?
  - Bandwidth?
- Their importance is relative and depends on the application.

Latency may Dominate Bandwidth: A Keystroke

- Consider the following application:
  - A client sends a keystroke (this is 1 Byte) to a server and receives back a 1 Byte message.
- The application performs differently across the room than across the transcontinental channel.
  - Whether the channel is 1 $Mbps$ or 100 $Mbps$ is not very relevant:
    (a) in a 1 $Mbps$ channel this takes 8 μs, while
    (b) in a 100 $Mbps$ channel it takes 0.08 μs.
- Here bandwidth is insignificant!

Bandwidth may Dominate Latency: An Archive

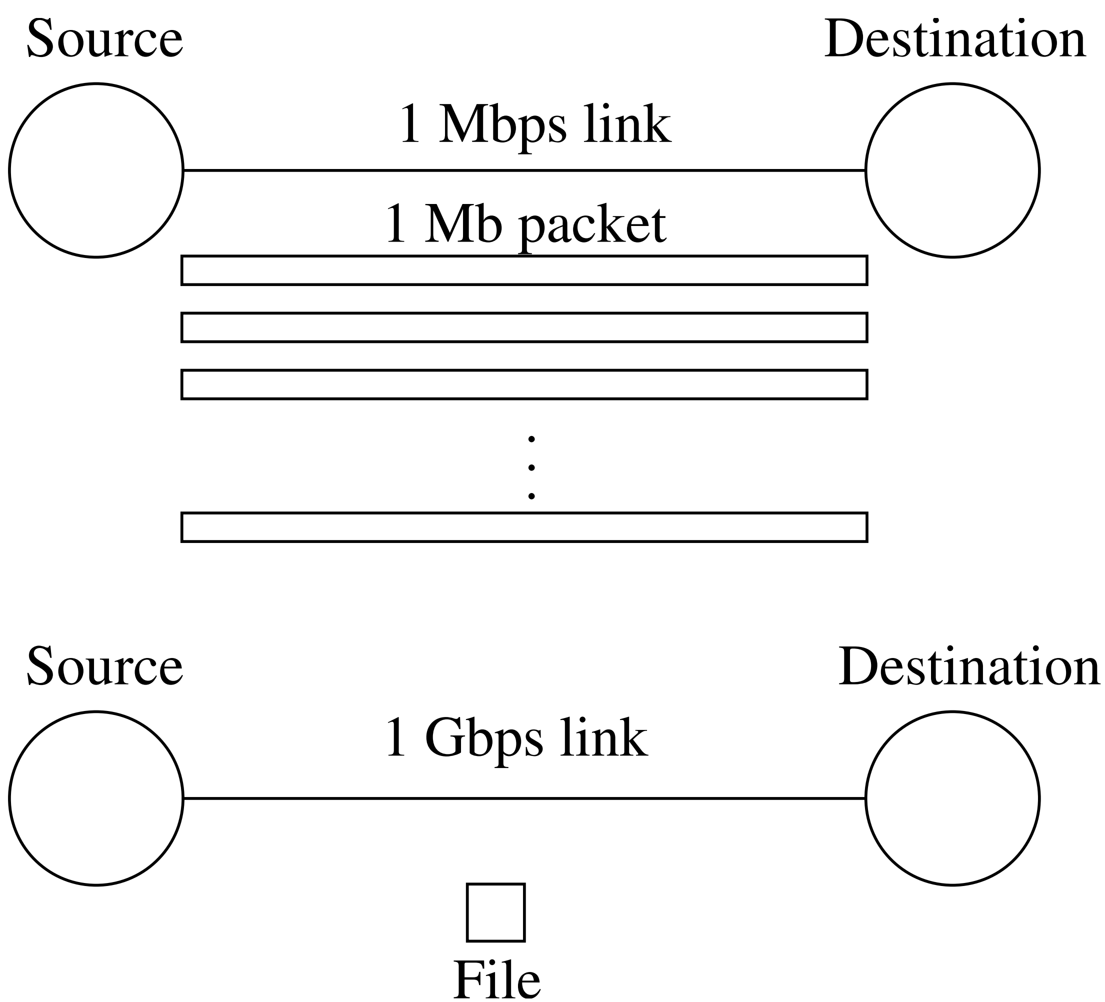
- Consider the following application:
  - A client is downloading a 25 MB archive from a digital library.
- In this case the more the bandwidth available the faster it will return the archive to the client.
- It takes

$$\frac{25\ MB}{10\ Mbs} = \frac{200\ Mb}{10\ Mbs} = 20\ sec$$

  to transmit this in a channel with bandwidth 10 $Mbps$.
- It is not very important now on whether the archive is located across the room or across the continent.
  - In the former case (a) the response time is 20.001 sec and in the latter case (b) 20.1 sec.

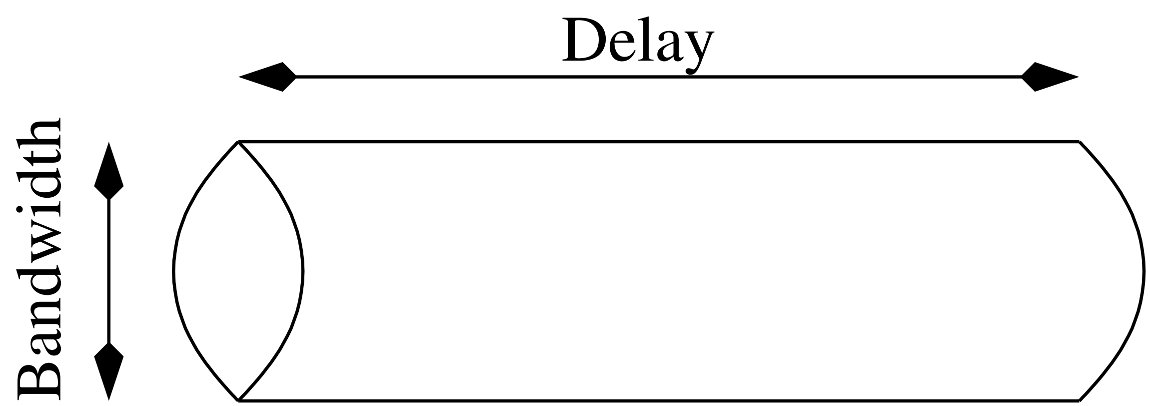
Impact of Speed: An Intuitive Comparison

A 1 MB file over a 1 Mbps- and 1 Gbps-network.



"Delay $\times$ Bandwidth" Product

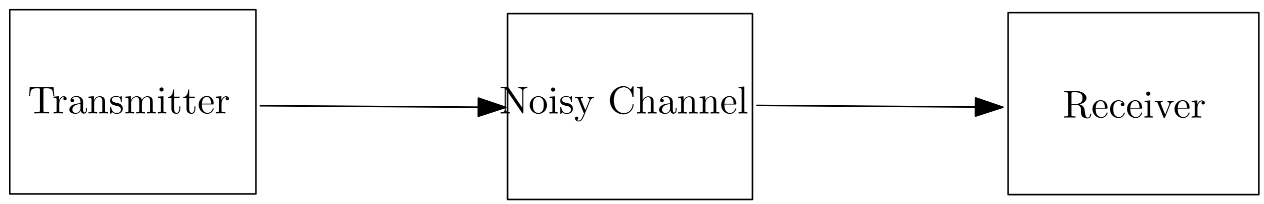
- Delay is measured in "time units"
- Bandwidth in "bits per time-unit".
- The Delay-Bandwidth Product is measured in bits.



- Delay-Bandwidth Product corresponds to the number of bits the sender must send before the first bit arrives at the receiver.
- Also the sender will send $2 \cdot (\text{Delay} \times \text{Bandwidth})$ bits of data before it receives a bit from the receiver!

Channel Capacity

- Data are transmitted through channels



...which can be imperfect!

- Quality affected by
  - bandwidth,
  - signal strength,
  - noise level.
- Let S = be the signal strength and N = noise level.
  - This gives rise to the Signal to Noise ratio $R := \frac{S}{N}$.
- Let the bandwidth be B.

Shannon Capacity

- What is the maximum capacity in bits per sec of the channel?
- **Shannon Capacity** C measures precisely this theoretically highest maximum:

$$C = B \log\left(1 + \frac{S}{N}\right) \text{ in bits per sec.}$$

- This is a mathematically sophisticated result!
  - You may wonder where does the log come from?
- An extremely noisy channel will have Signal to Noise ratio R close to "zero":
  - this is because $\log(1 + R)$, and hence also the capacity C is close to "zero".

Shannon Capacity: Example

- A typical telephone line has bandwidth $3,000\ Hz$ and signal to noise ratio of $SN = 3,162$ (usually measured in dB).
- Hence,

$$\begin{aligned} C &= B\log(1+SN) \\ &= 3000 \times \log(1+3162) \\ &= 34860\ bps \end{aligned}$$

- Hence, 34860 *bps* is the highest bit rate of a telephone line.

Information

- Justification for Shannon's capacity is based on a "very basic" but "subtle" question concerning "information".
- Given a random experiment with n equiprobable outcomes $E_1, E_2, \ldots, E_n$ (each may happen with probability p) how much information is conveyed on the average by a message M telling us which of these events occurred?
- A reasonable measure of this would be "the length of the message M" .... however written in an economical way!
  - Coin (H, T): one bit $0, 1$
  - Dice $(1, 2, 3, 4, 5, 6)$: three bits $000, 001, 010, 011, 100, 101$.
- Each outcome E_i will be encoded as a bit sequence in binary

$$b_1^i b_2^i \cdots b_\ell^i$$

Example

- Consider the outcomes of rolling a die:
  - each side occurring with probability $p = 1/6$.
- Such a sequence of outcomes would be $E_1, E_2, \ldots, E_6$

$$\text{“1”}, \text{“2”}, \text{“3”}, \text{“4”}, \text{“5”}, \text{“6”}$$

- Observe that we have $n = \frac{1}{p} = 6$ possible outcomes!
- Notice that you need 3 bits to encode the outcome, and
  - 3 is the smallest exponent ℓ such that $6 \leq 2^\ell$.

Shannon Capacity and Information

- Observe that in the general setting you have $n = \left\lceil \frac{1}{p} \right\rceil$ events.
- If we use binary codes of length ℓ to encode this message we would need to choose ℓ so that $n \leq 2^\ell$.
- The min such value of ℓ must satisfy

$$0 \leq \ell - \log_2 n < 1$$

  Why is the above true?
- The quantity

$$I := \log_2 \left(\frac{1}{p} \right) (= \log_2 n)$$

  is a reasonable measure of the average amount of information in the message M telling us which of these events occurred.

Explanations on Shannon Capacity: About $\log_2(1+\frac{S}{N})$

Assume a channel of unit bandwidth: $B = 1$.

- For $N = noise, S = signal$ uniformly random, $p = \frac{N}{S+N}$ will be the probability of "noise".
- Now look at the logarithm

$$\log_2\left(1+\frac{S}{N}\right) = \log_2\frac{S+N}{N} = \log_2\frac{1}{\frac{N}{S+N}} = \log_2\left(\frac{1}{p}\right).$$

- Set $p := \frac{N}{S+N}$: assume $N = 1$
  - if $S = 1$ then $p = 1/2$
  - if $S = 3$ then $p = 1/4$
  - if $S = 2^k - 1$ then $p = 1/2^k$
- So, $\log_2\left(1+\frac{S}{N}\right)$ represents the average amount of "bits" (or information) per unit bandwidth that can be transmitted in such a channel.

UNIX ping command

- **ping**
  utilizes the ICMP protocol's ECHO_REQUEST datagram to elicit an ICMP ECHO_RESPONSE from the specified host or network gateway.
- ping -s hostname
  When the -s flag is specified, ping sends one datagram per second

ICMP

- **ICMP** (Internet Control Message Protocol) is the error and control message protocol used by the Internet protocol family. It is used by the kernel to handle and report errors in protocol processing.
- ICMP is a datagram protocol layered above IP. It is used internally by the protocol code for various purposes including routing, fault isolation, and congestion control.
- Receipt of an ICMP "redirec" message will add a new entry in the routing table, or modify an existing one.

Exercises[a]

1. The performance of a client-server system is influenced by two network factors: the bandwidth of the network and the latency Give an example of a network that exhibits high bandwidth and high latency. Then give an example of one with low bandwidth and low latency.
2. Besides bandwidth and latency, what other parameter is needed to give a good characterization of the quality of service offered by a network used for digitized voice traffic?
3. List four different types of delay.
4. Calculating the frequency of a repeating event is accomplished by counting the number of times that event occurs within a specific time period, then dividing the count by the length of the time period. For example, what is the frequency if 91

[a] Not to hand in! A ($\star$) indicates the exercise may be harder than usual.

events occur within 15 seconds?

5. Edholm's law predicts that the bandwidth and data rates double every 18 months, which has proven to be true since the 1970s. Thi is evident for Internet, cellular (mobile), wireless LAN, and wireless personal area networks.
   (a) Why the more the bandwidth of an application the more data you can transmit?
   (b) When the bandwidth is large in wireless communications it takes longer for devices to discover each other? Why?
   (c) How can we improve performance of devise discovery when the bandwidth is large?
6. Think of the bandwidth B as length.
   (a) A device is scanning the bandwidth with speed v. How long does it take to scan the entire bandwidth?
   (b) ($\star$) Assume the bandwidth B is divided into m equal length

sub-bands each of size B/m. Two users U_1, U_2 pick one of these parts independently and uniformly at random. What is the probability they will discover each other?

(c) ($\star$) One of the two users U_1 has already selected a sub-band but this is unknown to the other user U_2. Assume that scanning is done with speed v. How long does it take in expectation for U_2 to discover U_1?

7. An application supports a frequency band ranging from 3 KHz to 3.7 KHz. What is Its bandwidth in Hz.

8. A client-server system uses a satellite network, with the satellite at a height of 40,000 km. What is the best-case delay in response to a request?

9. List advantages and disadvantages of a broadcast subnet when in use by multiple hosts.

10. ($\star$) A group of $2^n - 1$ routers are interconnected in a

centralized binary tree, with a router at each tree node. Router i communicates with router j by sending a message to the root of the tree. The root then sends the message back down to j. Derive an approximate expression for the mean number of hops per message for large n, assuming that all router pairs are equally likely.

11. ($\star$) In some networks, the data link layer handles transmission errors by requesting damaged frames to be retransmitted. If the probability of a frame being damaged is p, what is the mean number of transmissions required to send a frame? Assume that acknowledgements are never lost.