The Physical Layer

Mechanisms for encoding and transmitting information

- Varying voltage: Electrical conductor
- Varying RF signal: Space
- Varying optical signal: Optical fiber

Elements of a Signaling Protocol Specification

Definition of signal states (Physical state X corresponds to bit pattern Y)

N distinct signal states ==> \( \log_2(N) \) bits per signal state

Consider four color flashlight signaling

- Red: 00
- Green: 01
- Blue: 10
- White: 11

Duration of signaling states \((1 / \text{Duration} = \text{Baud Rate})\)

- \(\text{bits/baud} \times \text{baud rate} = \text{bit rate}\)
- \(\log_2(\text{N}) \times 1 / \text{Duration} = \text{bit rate}\)

Clocking

How to synchronize receiver and sender with respect to changes in the state of the signal (periodic resynch or self-clocking).
Example of a base band signal

+12v
+0v

Clock + + + + + + +
Data 0 1 0 0 1 0
Question: How fast can data be sent on a given channel?

Answer: There are two constraints

Bandwidth limitations (the Nyquist limit)

constrain how frequently you may change signaling states and thus defines a maximum useful baud or signalling rate.

Signal to noise limitations (the Shannon limit)

constrain the number of bits per signaling state (bits per baud) that may be encoded. As the channel becomes more noisy, it becomes impossible to distinguish reliably states that are too close to each other. And with a fixed power budget its not possible to create states that are not too close to each other.
Fourier series representation

Provides best mechanism for understanding the Nyquist limit.

Any (reasonable) function \( f(t) \) defined on an interval \([0, T]\) can be represented as:

\[
f(t) = \frac{c}{2} + \sum_{n=1}^{\infty} \left( a_n \sin \left( \frac{2 \pi n t}{T} \right) + b_n \cos \left( \frac{2 \pi n t}{T} \right) \right)
\]

where

\[
a_n = \frac{2}{T} \int_{0}^{T} f(t) \sin \left( \frac{2 \pi n t}{T} \right) dt
\]

\[
b_n = \frac{2}{T} \int_{0}^{T} f(t) \cos \left( \frac{2 \pi n t}{T} \right) dt
\]

\[
c = \frac{2}{T} \int_{0}^{T} f(t) dt
\]
Harmonics and Bandwidth.

The terms \( \sin(2 \pi n \frac{t}{T}) \) and \( \cos(2 \pi n \frac{t}{T}) \) are called the \( n^{th} \) harmonics of the series. These are sinusoidal functions that complete \( n \) full-period oscillations in time \( T \).

Transmission media tend to pass harmonics only within a certain finite range. Harmonics lying outside this range are strongly attenuated (reduced) in power.

For example, an AM radio signal is not strongly attenuated by a brick wall but visible light is.

The number of harmonics passed is called the bandwidth of the medium. Artificial bandwidth constraints may also be imposed with filters.

If a signal is passed through a medium that passes harmonics

\[ [J + 1, K] \text{ then the } \textit{bandwidth} = K - J, \]

The output includes harmonics \( J + 1 \leq n \leq K \) and can be represented as the Fourier series

\[
\hat{f}(t) = \sum_{n=J+1}^{K} (a_n \sin(2 \pi nt/T) + b_n \cos(2 \pi nt/T))
\]

Therefore, the input signaling function \( f(t) \) should not contain harmonics outside the range \( [J + 1, K] \) because they will be \textit{completely eliminated by the filtering effect of the medium}. 
The Nyquist Limit

Suppose we sample the function \( \hat{f}(t) \) at a rate fast enough to collect \( 2(K-J) \) samples in the interval \([0, T]\). Call the values \( f_1, f_2, \ldots, f_{2(K-J)} \).

Then we could (theoretically) set up the linear system with the \( a \)'s and \( b \)'s being the unknowns, the \( t \)'s being the known times at which we measured the function and the \( f \)'s being the measured values of the functions.

\[
\begin{align*}
    f_1 &= \sum_{n=(J+1)}^{K} (a_n \sin(2\pi nt_1/T + b_n \cos(2\pi nt_1/T))) \\
    f_2 &= \sum_{n=(J+1)}^{K} (a_n \sin(2\pi nt_2/T + b_n \cos(2\pi nt_2/T))) \\
    \vdots & \\
    f_{2(K-J)} &= \sum_{n=(J+1)}^{K} (a_n \sin(2\pi nt_{2(K-J)}/T + b_n \cos(2\pi nt_{2(K-J)}/T)))
\end{align*}
\]

Thus we have \( 2(K-J) \) equations and \( 2(K-J) \) unknowns \( a_{J+1}, \ldots, a_K \) and \( b_{J+1}, \ldots, b_K \).

If we solve this system we recover the exact signaling function.

In reality it is far simpler and more efficient to recover the coefficients via finite fourier transform.

Therefore, since sampling the signal at rate \( 2(K-J) \) allows us to exactly reconstruct the signaling function, sampling at rates faster than twice the bandwidth cannot possibly yield additional information.

Thus the Nyquist Theorem says:
The maximum useful baud rate is \( 2 \times \) bandwidth. Combining this with our previous result we see:

\[
\begin{align*}
    H &= \text{bandwidth} \\
    V &= \text{the number of states of the signal} \\
    \text{Max data rate} &= 2H \log_2(V)
\end{align*}
\]
The Nyquist result says nothing about limiting V

V might represent the number of different voltage levels recognized on a base band signal.. [0, 12], or [0, 4, 8, 12], or [0, 2, 4, 6, 8, 10, 12],.... etc.

How far can they be subdivided before levels become indistinguishable???

The answer, due to Shannon, is that it depends upon the ratio of the signal power to the power of the thermal (and other externally induced) noise in the carrier.

Shannon's result:

The Signal to Noise ration (SNR) on a channel is typically expressed in decibels as

\[ SNR = 10 \log_{10} \left( \frac{S.P.}{N.P.} \right) dB. \]

However, the Shannon limit uses the base 2 logarithm actual ratio of signal power to noise power.

\[ Max \ data\ rate = H \log_2 (1 + \frac{S.P.}{N.P}) \]

where H is again the Bandwidth. The first step in computing the maximum data rate of a given channel is to convert the SNR in dB to S.P. / N. P.
Example (Voice phone line):

\[ H = 4000 \text{ HZ} \]
\[ \text{SNR} = 30\text{DB} = 10 \log_{10} \left( \frac{\text{S.P.}}{\text{N.P.}} \right) \]
\[ \Rightarrow \log_{10} \left( \frac{\text{S.P.}}{\text{N.P.}} \right) = 3 \quad \Rightarrow \quad \frac{\text{S.P.}}{\text{N.P.}} = 1000 \]

Max data rate = \( 4000 \log_2(1001) \approx 40000 \text{ bps.} \)
(In practice very difficult to achieve!)

By Nyquist's Theorem the maximum sampling rate is
\[ 2 \times 4000 \text{ Hz} = 8000 \text{ Hz} \]

To obtain 40000 bps we required \( \frac{40000}{8000} = 5 \text{ bits per sample} \)

Thus, since \( \log_2(32) = 5 \), a \underline{32 state} encoding must be used.
**Analog communication links**

Historical basis for telephone, radio, and television.
Simple and cheap to construct
The signal carried on the wire is the analog of some physical process (e.g. the air pressure differentials that produce audible sound)

Telephone principles:

Audible sound range is approximately (30-20000)Hz.
Microphone: An electrical voltage is generated which is the analog of the varying pressure differentials created by voice.
Speaker: Varying air pressure differentials are regenerated.
Telephone system passes harmonics from 100 - 4000 hz.

Transmitting digital data on an analog carrier

Bit Serial (serial port) input / output of a computer is baseband signaling (RS-232 typically), but a *modulated sine wave carrier* rather than baseband signaling is used over the phone systems and other carriers.

The general sine wave carrier has form

\[ f(t) = A \sin(wt + q) \]

Information is encoded by varying one or more of the parameters as a function of *time*

- \( A(t) \) Amplitude modulation
- \( w(t) \) Frequency modulation
- \( q(t) \) Phase

A modem encodes a base band signal and decodes a modulated signal.
Historical Standard modems

Bell 103 - 300 bps at 300 baud FM (FSK)

<table>
<thead>
<tr>
<th></th>
<th>Mark (1)</th>
<th>Space (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Originate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xmit</td>
<td>1270</td>
<td>1070</td>
</tr>
<tr>
<td>Recv</td>
<td>2225</td>
<td>2025</td>
</tr>
</tbody>
</table>

Answer

<table>
<thead>
<tr>
<th></th>
<th>Mark(1)</th>
<th>Space(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xmit</td>
<td>2225</td>
<td>2025</td>
</tr>
<tr>
<td>Recv</td>
<td>1270</td>
<td>1070</td>
</tr>
</tbody>
</table>

Bell 212 - 1200 bps at 600 baud PM (PSK)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Originate</td>
<td>??</td>
<td>45 = 00</td>
</tr>
<tr>
<td>Xmit</td>
<td>***</td>
<td>135 = 01</td>
</tr>
<tr>
<td>Recv</td>
<td>**</td>
<td>225 = 10</td>
</tr>
<tr>
<td>Answer</td>
<td>315 = 11</td>
<td></td>
</tr>
<tr>
<td>Xmit</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Recv</td>
<td>??</td>
<td></td>
</tr>
</tbody>
</table>

Bell 203 - 9600 bps at 2400 baud (Combined AM/ PM)

12 Phase changes with two amplitudes at each of 4 phases
Quadrature Amplitude Modulation

Modern modems use amplitude modulation on two phase shifted carriers. This technique is called QAM-n (Quadrature Amplitude Modulation) where \( n \) is the number of distinct signaling states and is often called the number of \textit{symbols}.

With \( N \) symbols \( \log_2(N) \) bits per baud may be transmitted.

QAM employs two sinusoidal carriers

- The I channel at nominal phase shift 0
- The Q channel at phase 90 degrees.

Properties of the sin and cos function make it possible to additively combine the two carriers at the sender and then recover the original signals at the receiver.
QAM transmitter

QAM receiver

Note: This diagram was downloaded from somewhere in the WorldWideWeb. I make no claims of authorship... Dittos for next page.
A QAM-16 constellation

In QAM-16 four distinct amplitudes are used on both the Q and I channels. This allows use to encode two bits on the Q and two bits on the I channel each symbol time. The result is 16 possible symbols as shown. In each quadbit group \textit{the first two bits are the I channel values}. Proceeding left to right we have 00, 01, 11, 10. The \textit{last two bits are the Q channel values}. Proceeding top to bottom the same layout occurs.

Note that by splitting the stream, we can get 16 symbols or 4 bits per baud with only 4 different amplitude levels.
The QAM – 64 constellation is shown in the book. It is an 8 x 8 square analogous to the one above. Since \( \log_2(64) = 6 \) the sixty four symbols can encode 6 bits per baud, but only 8 distinct amplitude levels are required instead of the 64 that would be required for pure amplitude modulation.

The spectral efficiency of QAM can be seen in the following table when compared to pure AM can be seen in the following table

<table>
<thead>
<tr>
<th>QAM-N</th>
<th>#Symbols</th>
<th>#Amplitudes</th>
<th>#AM amplitudes</th>
<th>Bits/Baud</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>4</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>64</td>
<td>64</td>
<td>8</td>
<td>64</td>
<td>6</td>
</tr>
<tr>
<td>256</td>
<td>256</td>
<td>16</td>
<td>256</td>
<td>8</td>
</tr>
</tbody>
</table>

With pure amplitude modulation the number of amplitudes is \( 2^n \) where \( n \) is the number of bits per baud.

With QAM the number of amplitudes is \( \sqrt{2^n} = 2^{n/2} \)

Commercial modems using QAM include:

- V.32bis 7 bits/symbol – 6 data + 1 parity
  2400 baud
  14,400 bps

- V.34bis
  14 bits / symbol
  2400 baud

QAM is also widely used in wireless and cable systems. In these adaptive modulation approaches are common. The encoders can use QAM 4, 16, 64, or 256 depending upon the ambient SNR.
Digital communication channels

Disadvantages of traditional analog transmission circuits (those that can carry analog voice and video transmissions)

Analog signals have are arbitrary wave patterns and thus can't be regenerated but must be amplified. Amplification amplifies NOISE and inevitably destroys the signal.

A digital communication channel carries a signal that always exists in one of a small number of states (often two states) and changes states only at fixed time intervals.

A baseband signal is an example of a digital signal BUT NOT ALL DIGITAL SIGNALS ARE BASEBAND.

A digital signal can be regenerated instead of amplified since it always exists in a small number of states ==> It can be transmitted any distance with no loss of quality.

A digital signal can be transmitted and regenerated on an analog channel but regeneration will destroy any analog signal.

Thus a digital signal is ideal for reliable high speed data service and high speed data services use exclusively digital channels.
Carrying analog telephone signals on a digital channel

An analog signal is encoded using *pulse code modulation*.

From the Nyquist theorem we know that to reproduce the signal exactly we must sample at twice the highest frequency we wish to capture.

Analog telephones carried up to 4000 Hz.

So analog voltage is sampled and digitized every 125 microseconds (8000 / times per second) using an A-D converter.

Voltage is encoded as one of 256 discrete values

The 8 bit sample is sent to the receiver where...

an output voltage equivalent to the input is generated by D-A converter.

Bandwidth requirements:

8000 samples / second * 8 bits per sample = *64000 bits per second.*

This is the fundamental unit of telephone channel capacity..

It is sometimes called a *DS-0.*

Higher fidelity sound systems require higher bit rates

CD Audio

44,100 samples / second => harmonics up to 22 KHz may be encoded

16 bits (or higher) per sample => higher resolution encoding

Total bit rate = *705600 bps per channel.*
Multiplexing

Superimposing multiple logical connections on a single physical channel. Three generations of multiplexing have been used in the telephone system.

Original Motivation:
Cheaper to run a single high bandwidth cable that 100,000 voice grade links between two large cities!

Two common strategies:
- **FDM**: Frequency division multiplexing
- **TDM**: Time division multiplexing

Generation one: FDM -
Used in (VERY obsolete) analog long distance telephone circuits.

- e.g. telephone groups
  - Group: 12 4 KHz channels between 60 and 108 Khz
  - Supergroup: 5 groups
  - Mastergroup 10 supergroups
Generation 2: Synchronous TDM -

Used on digital interoffice wired trunks
Local loop POTS is still analog
No guard bands are required.
Channel capacity is broken down into logical slots.
Slot size is the number of bits consumed at a time from each individual channel
Slots are assigned round robin
A set of slots from every channel is called a frame

<table>
<thead>
<tr>
<th>D3</th>
<th>D2</th>
<th>D1</th>
<th>D0</th>
<th>D3</th>
<th>D2</th>
<th>D1</th>
<th>D0</th>
</tr>
</thead>
<tbody>
<tr>
<td>slot</td>
<td>slot</td>
<td>slot</td>
<td>slot</td>
<td>slot</td>
<td>slot</td>
<td>slot</td>
<td>slot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>frame</td>
<td></td>
<td>frame</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TDM in the early digital (electrical) telephone network

T1 (DS1)
slot size = 8 bits
24 PCM (DS0) + 1 signaling bit = 193 bits per frame
| ch0 | ch1 | ch2 | .......... | ch23 | synch |
8000 frames / second x 193 bits / frame = 1.544 Mbps

T2 (DS2)
bit interleaved muxing of 4 T1’s
management overhead => aggregate bit rate of 6.312 Mbps

T3 (DS3)
bit interleave muxing of 6 T2’s
more management over raw bit rate = 44.736 Mbps
effective rate = 24 * 1.544 = 37.06 Mbps
or roughly 17% overhead!

At this speed electrical signaling over copper channels “hit the wall” and a new technology was built from the ground up for the current system.
Generation 3: Optical TDM in the modern digital telephone system

Sonet/SDH (OC-1 (optical carrier) or STS-1 (synchronous transport signal))

Frame = a 9 row x 90 column BYTE matrix

8000 frames / second = 51.84 Mb/second (the 125usec inter-frame time is required for the standard 8Khz PCM voice channel... Each PCM call gets one byte / frame)

First 3 columns is reserved for system management overhead
These columns are allocated as
First 3 rows = section overhead
Next 6 rows = line overhead
Total Overhead = 3 / 90

User data = 87 x 9 x 8 x 8000 bps = 50.112Mbps

Payload can begin anywhere and span frames
1st column of Payload = path overhead

OC-n 3,9, 12, 18, 24, 36, 48 is n x an OC-1

Multiplexing of STS channels is done at the byte level.

The term OC-nC is means OC-n catenated. It is used to describe channels in which the full capacity is dedicated to a specific use such as a high speed data channel (instead of multiplexed telephone traffic).
**Other multiplexing mechanisms**

WDM (Wave division multiplexing) -

A bleeding edge FDM technology used on optical links.

A single fiber may have bandwidth exceeding 25,000 GHz. A 10 bits / hz modulation scheme could thus carry 250 trillion bits per second on a single fiber!

Using a single frequency (\(\lambda\)) it is now common to operate a single fiber at OC-192 (10 Gbps).

However, it is now possible for up to 96 different \(\lambda\)s to coexist on a single fiber providing a per-fiber long haul capacity of 960 Gbps!

Electrical to optical conversion is limited to a few GHz

A prism or diffraction grating can be used in ADM's (Add/Drop Multiplexors) and in switching to avoid all electrical conversions on the optical superhighway.
ATDM (asynchronous TDM a.k.a. concentration)

Distribute all (or unused slots) on a demand basis
Requires addressing overhead but...
Overhead is compensated for by lack of wasted slots if the demand is bursty

|   D4 - A4 |   D2 - A2 |   D4 - A4 | ..... |
| slot      | slot     | slot     |

ATM is an example of ATDM

For STDM slot size has no impact upon efficiency, but for ATDM larger slot size implies higher efficiency, but too large may produce undesired latency for QoS sensitive traffic due to head of line blocking.

Hybrid TDM

Multiplex ATM channels over SONET links
ATM -

ATM employs connection oriented, ATDM cell switching based technology

A cell is a small (53 byte) fixed length packet 48 + 5 that can contain
   48 PCM slots
   48 bytes of data

Recall that minimizing cell size produces minimal latency in hybrid switched networks.

Designed for heterogeneous loads
   Video
   Voice
   Image
   Data

Designed to support high data rates
   100 Mbit + links
   10 Gbit + switches

Cell units managed by switches on a store and forward basis
Priority based services possible
ISDN - Integrated services digital network  (a.k.a It Still Does Nothing)

Essential concepts... a collection of well defined STDM channels

Channels:
  B - 64kps PCM for voice or data
  D - 16kpbs for signaling

Services
  Basic rate 2B+1D
  Primary rate 23B + 1D (almost a T1).

Providing digital bit pipe services
  Connection oriented
  Bit stream
  unreliable

Call connection => full bandwidth reserved for that connection.

Since each line consisted of purely digital channels codecs were required for voice
ADSL - Asymmetric digital subscriber line

Essentially an admission of the failure of ISDN
Requires removing filters from the local loop
Limited in distance from local office or your neighborhood concentrator (up to 5.5km ... but your mileage may vary)
Asymmetric in that more downstream bandwidth is available than upstream

Characteristics of original AT&T service:

- Low 25Khz of the circuit is reserved for voice (only 4 Khz is still used)
- Other 21Khz is a buffer to reduce crosstalk
- Remainder of the bandwidth is either...
  - FDM multiplexed between upstream and downstream service
    - Upstream 25Khz - 200Khz
    - Downstream 250Khz - 1Mhz

- or Echo cancellation is used to make full bandwidth available both ways
Modern DSL systems use DMT (Discrete Multitone) Signaling

DMT is a form of OFDM (orthogonal frequency division multiplexing).

In OFDM the entire bandwidth of the channel is subdivided into subchannels having mutually orthogonal carrier frequencies. OFDM is commonly used with QAM subchannels. (Recall that the individual harmonics of the Fourier transform are mutually orthogonal).

\[ f(t) = \frac{c}{2} + \sum_{n=1}^{K} (a_n \sin(2\pi nt/T) + b_n \cos(2\pi nt/T)) \]

*Encoding the signal:*

The number of channels is \(2^n\) for some \(n\).
An input signal splitter partitions the signal among \(m < n\) modulators
Each channel uses a modulation scheme producing a signal \(X_j(t)\)
The *inverse* Fast Fourier transform is used to combine the \(X_j\)'s

*Decoding the signal:*

The FFT is used to split the combined signal in to the individual \(X_j\)'s
These are fed to demodulators
The output of the demodulators is recombined to produce a single bit stream

Forward error correction is commonly also used in the process.
Details of the ADSL implementation

In ADSL the 1.1 Mhz spectrum is divided into 256 subchannels of 4312.5 each Hz using the OFDM technique (Orthgononal frequency division multiplexing).

Channel 0 - POTS
Channel 1 - 4 Guard bands
2 additional channels set aside for upstream and downstream control
Remainder of the channels theoretically available for data... quality of line and cross talk may restrict usage though.
Typical setup
  32 channels upstream
  remainder downstream
Per channel modulation is 4000 baud QAM.
End user ADSL modems usually export an ethernet interface.
Cable modem systems

Sometimes called HFC (hybrid fiber cable)
Built on fiber optic trunks with shared-medium copper coaxial cable to end users.

Upstream and downstream channels operate in different frequency ranges permitting simultaneous upstream and downstream traffic (along with TV signals).

A head-end device called the Cable Modem Terminating System (CMTS) controls access to the media.

Various flavors of QAM are commonly employed.
**WiMAX example**

The individual channel bandwidth for the 4.9GHz Public Safety profile is 5 Mhz. Upstream and downstream transmissions share the same channel, and thus operation is necessarily half-duplex. The default configuration of the M/A-COM equipment deployed at Clemson is a 50/50 upstream/downstream split with radios alternating between transmit and receive mode every 5 ms.

Single carrier, OFDM, and OFDMA multiplexing are all supported in WiMAX. For OFDM at 4.9 Ghz the subchannels are allocated as follows:

- 256 Total subchannels
- 8 "Pilot" channels used to establish/maintain physical layer synchronization
- 192 Data channels
- 56 Channels used as guard bands.

Each subchannel has a bandwidth of $5 \text{ Mhz} / 256 \text{ channels} = 19.531 \text{ KHz}$.

The WiMAX standard specifies that the symbol/sampling/baud rate should be $144 / 125$ of the channel bandwidth\(^1\) or $22.50 \text{ Ksym / sec}$. Thus the symbol (baud) time is the inverse of $22,500$ or $44.4444 \text{ usec/symbol}$.

The symbol (baud) time is known in the WiMAX standard as the basic FFT symbol duration because the FFT is used in demultiplexing QAM encoded OFDM signals. However, the standard defines possible cyclic prefix or guard intervals of $1/4, 1/8, 1/16$ and $1/32$ of the FFT symbol duration. The 4.9GHz profile uses a $1/8$ symbol time guard duration which makes the effective symbol time $50 \text{ usec/symbol}$ or $20,000$ symbols per second per channel.

Since 192 channels carry data, with the minimum guard interval we obtain $192 * 20,000 \text{ Ksym} = 3.84 \text{ Msymbols / second}$ at the physical layer.

QAM-64 yields 6 bits per symbol or an aggregate physical layer bit rate of almost $23.04 \text{ Mbps}$. Note that the bit rate is almost five times the bandwidth. It might be noted that 1990's era modems could achieve a bit rate of 8 times the 4Khz bandwidth on voice grade telephone links, but SNR's encountered in wireless are also not nearly so good as those on even voice grade phone lines.

---

\(^1\) Recall that the *absolute* maximum useful baud rate is $2 \times$ the bandwidth.
In QAM-64 2/3, 1/3 of the total bits carried are reserved for ECC, yielding a data bit rate of 15.36 Mbps. For symmetrically provisioned systems this must be further reduced by a factor of 2 yielding a maximum of 7.68 Mbps. With a maximally sized guard interval this value is further reduced to 6.8 Mbps. Framing and other MAC level protocol overhead further reduce the application data bit rate.

For more information read wimax.pdf in the class directory.
Switching mechanisms

**Circuit switching** (historically the telephone model):

- Setup establishes a dedicated path between sender and receiver
- Data flows with no intermediate stops
- Major disadvantages:
  - Setup / takedown overhead
  - Wasted capacity for bursty traffic.

**Message switching** (historically the telegram model):

- Entire logical messages are transmitted on a store and forward basis.
- Disadvantages:
  - Must be able to buffer entire message at each switching center
  - End to end transmission time = # of hops * pt to pt transmission time.

**Packet switching** ("invented" in the ’60s)

- Logical messages are broken down into packets (having no connection to the semantics of the messages). Packets are limited to some fixed maximum size.
Message switching vs packet switching

In the old message switching model (based upon telegraphy) logical messages are transmitted through the network as continuous transmissions _regardless of the size of the message._

In packet switched networks messages are broken down into limited size packets, and the packets are transmitted as independent units on a store and forward basis. Two advantages of packet switching are:

- Packet switching can also be used to alleviate the _head-of-line_ blocking effect.

- Overlapped transmission of packets is possible and provides a pipelined parallel processing effect that can _reduce end to end latency and increase effective path throughput_ (when compared to message switching) _on multi-hop paths_. But the maximum path throughput in _bps_ is always limited to the throughput of the bottleneck link!

Consider the following example:

- 800000 bit message
- 10000 bits / second link layer throughput
- 10 hops
- Assume no queuing delays
Message switching:

Message transmission time = 800000 / 10000 = 80 secs
80 seconds / hop * 10 hops = 800 seconds

*Effective* throughput = 800000 / 800 = 1000 bits / second

Packet switching:

8000 bit packets
100 packets to send

Packet transmission time
8000 / 10000 = 0.8 secs
Last packet completes transmission at source at time 100 * 0.8 = 80 secs.
Last packet arrives at destination 9 * 0.8 = 7.2 seconds later
Total transmission time = 87.2 seconds

*Effective* throughput = 800000 / 87.2 = 9174 bits / second

Note the effective throughput is now *approaching* the physical layer bit rate

For zero overhead packetizing, the optimal packet size is *one bit!*
Cost of headers

In the real world packets must carry headers

Suppose \( H \) bit headers are used and packet size is \( P \) (including \( H \))

Then packet transmission time is \( P / 10000 \)

And the number of packets is \( 800000 / (P - H) \)

Time until the last packet leaves first hop is:

\[
\text{pkt-time} \times \# \text{of pkts} = \frac{P}{10000} \times \frac{800000}{(P-H)} = \frac{80P}{(P-H)}
\]

Additional time until last packet reaches destination is

\[
9 \times \frac{P}{10000}
\]

Total latency is thus

\[
\text{Latency} = \frac{80P}{(P-H)} + \frac{9P}{10000}
\]

For the minimum packet size of \( P = H + 1 \) the first term is dominant

\[
\text{Latency} = 80(H+1) + 9(H+1) / 10000
\]

For the maximum packet size of \( 800000 + H \)

\[
\text{Latency} \approx 80 + 9 \times 80
\]

which is approximately the same as message switching.