Sampling,
Multimedia Communications,
and Streaming

Rynson W.H. Lau

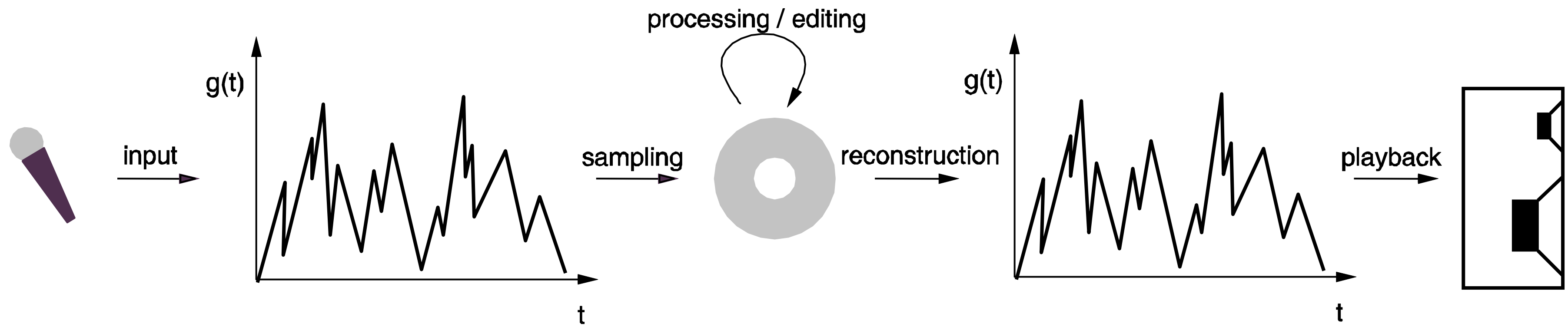
Key Topics

- Sampling
- Network Transmission
- Quality of Service (QoS)
- Wireless Networks
- Multimedia Streaming

Sampling

- Real world information are mostly continuous, meaning that you can significantly zoom in and look at the information.
- For example, when looking at a scenery (visual information), you can walk closer and closer to it and be able to see more and more details.
- On the other hand, our computers are digital in nature. They can only store/handle digital (or discretized) information.
- In order for the computer to store/process real world information, we need to convert the information from a continuous form to a digital form.

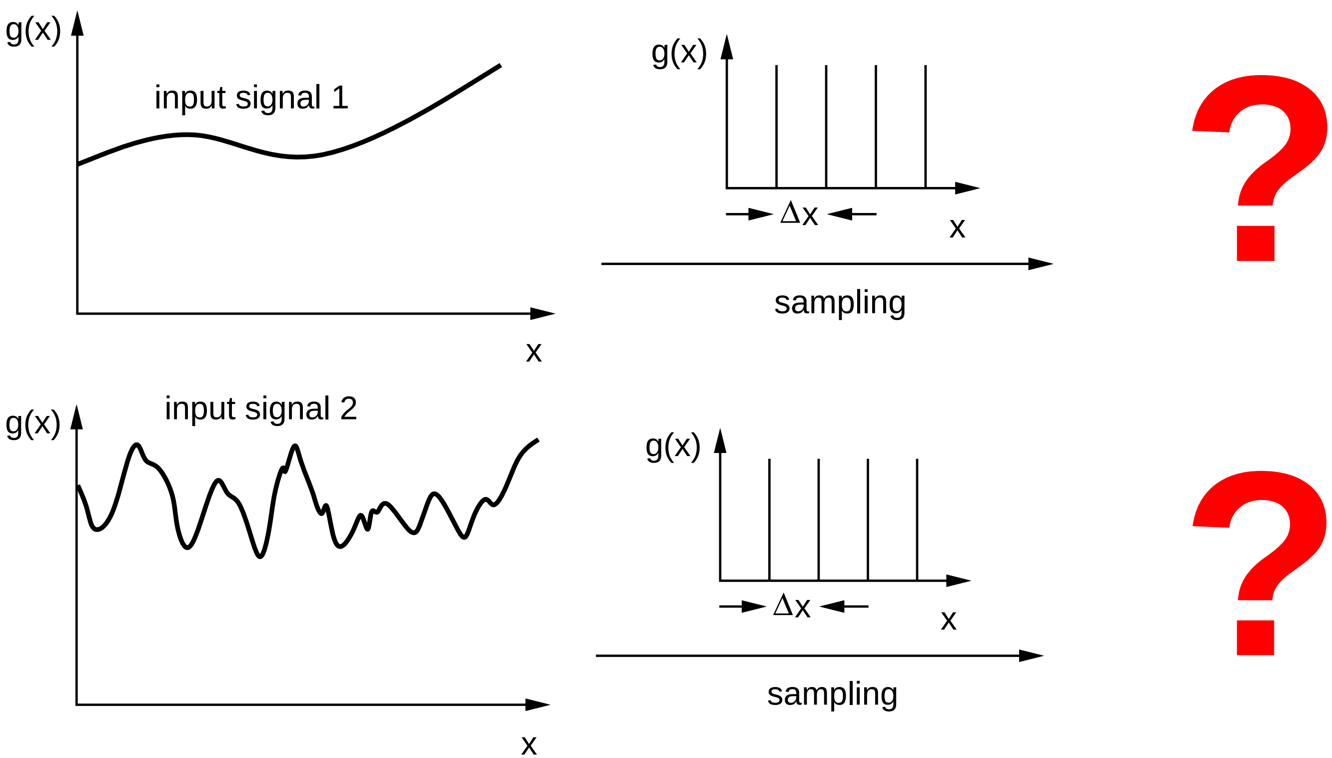
- ***Sampling*** is a process to convert a continuous signal, such as a voice signal, into digital form so that we may store / transmit / edit / playback the content.



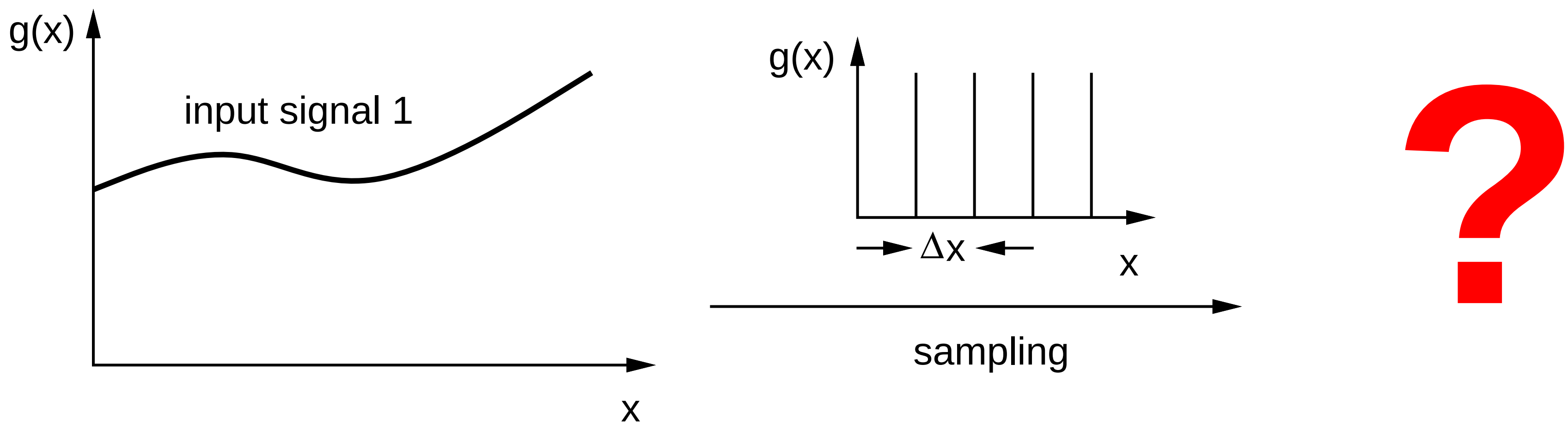
- Note that to the computer, the difference between audio and visual information is only that audio is 1D and visual is 2D.

- We want to know if a continuous signal, $g(t)$, can be sampled at some interval Δt without losing any information.
- For audio signals, $g(t)$ represents the input voltages from the microphone while Δt represents the sampling time interval.
- For images, $g(t)$ can be written as $g(x, y)$, which represents the gray-level (or color) values from the camera while Δx and Δy represent the distances between samples in x and y directions.

Consider two input audio signals, which are sampled and then reconstructed to form the output (playback) signals:

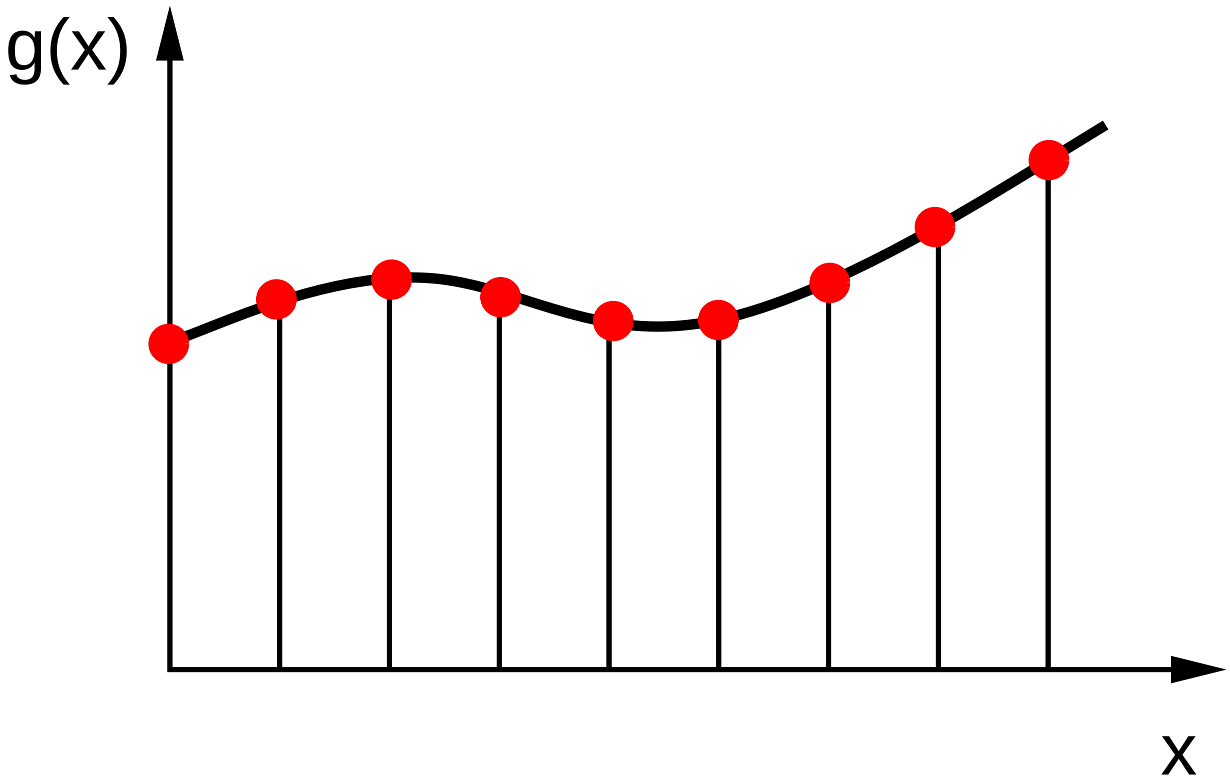


First, we look at the following signal, which changes much more slowly:

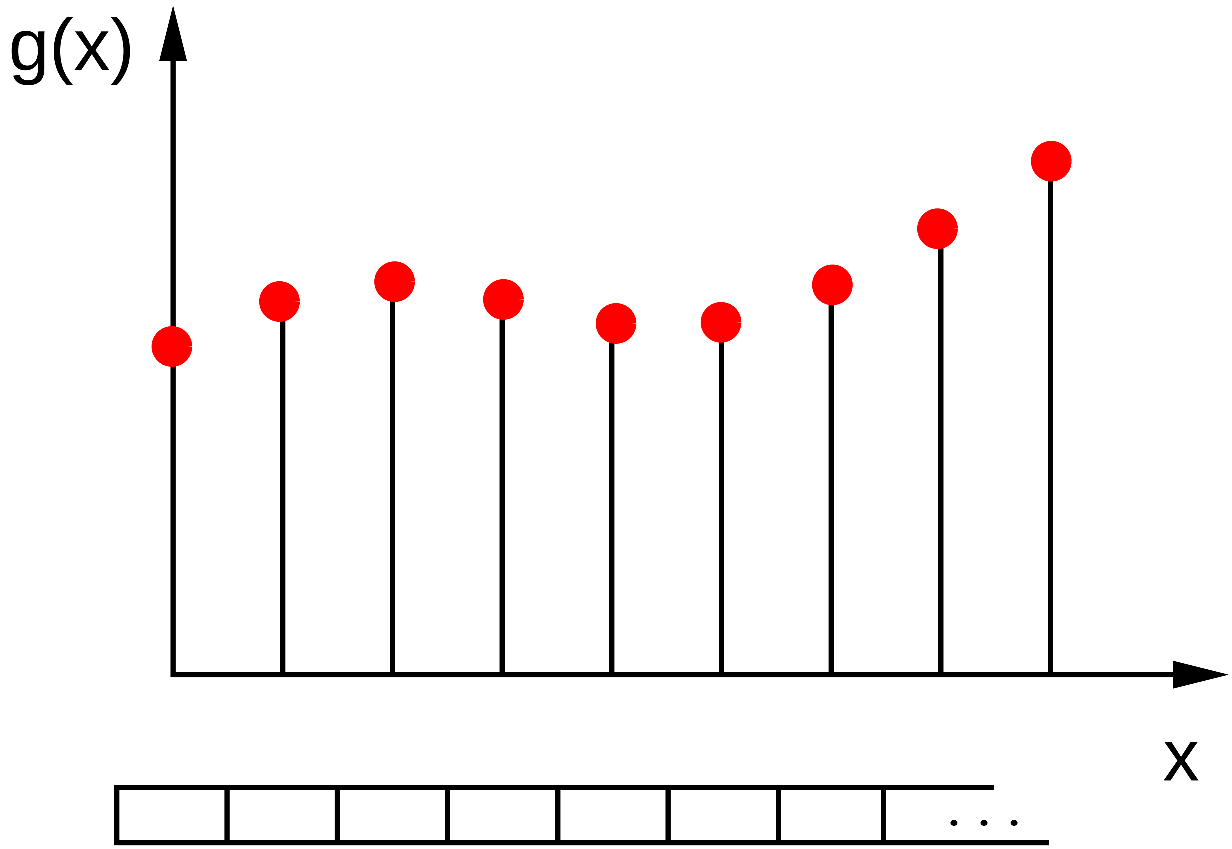



Original signal
g(x)
x

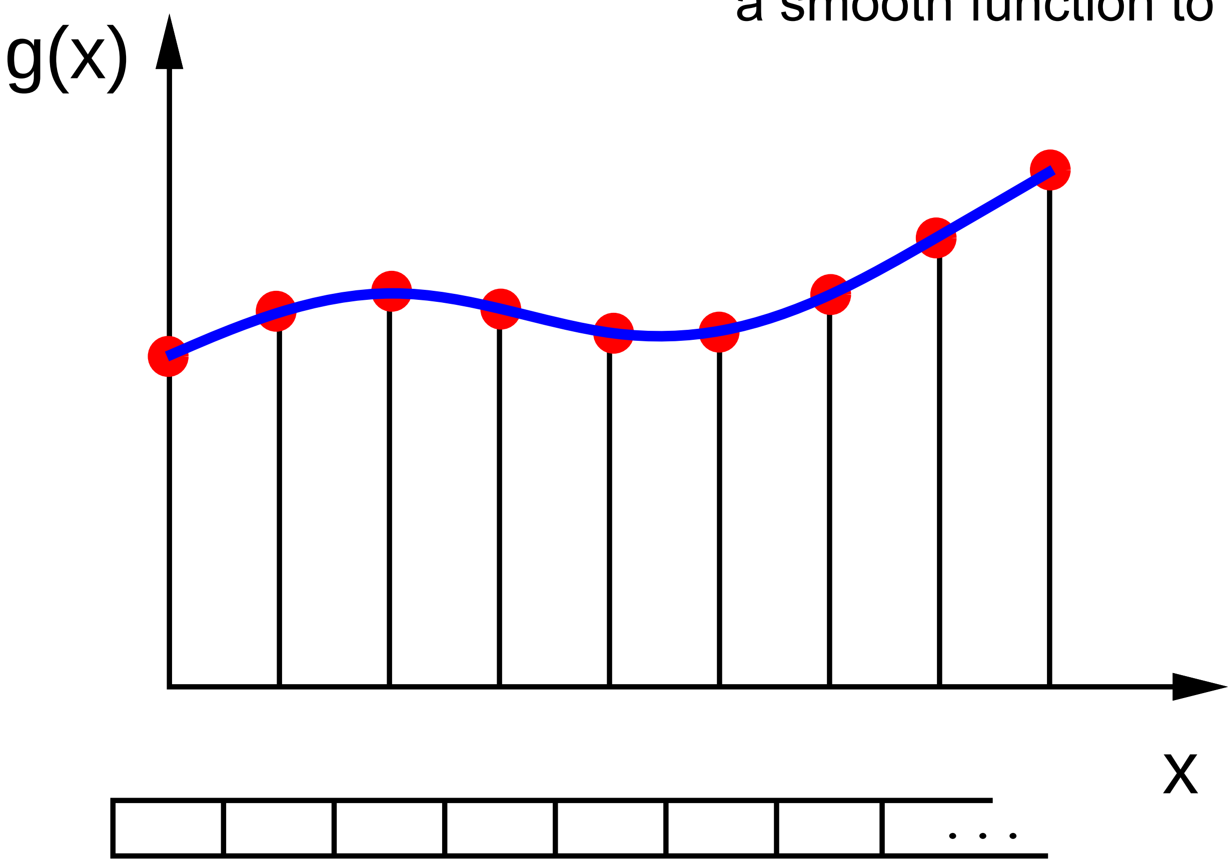
Original signal being sampled



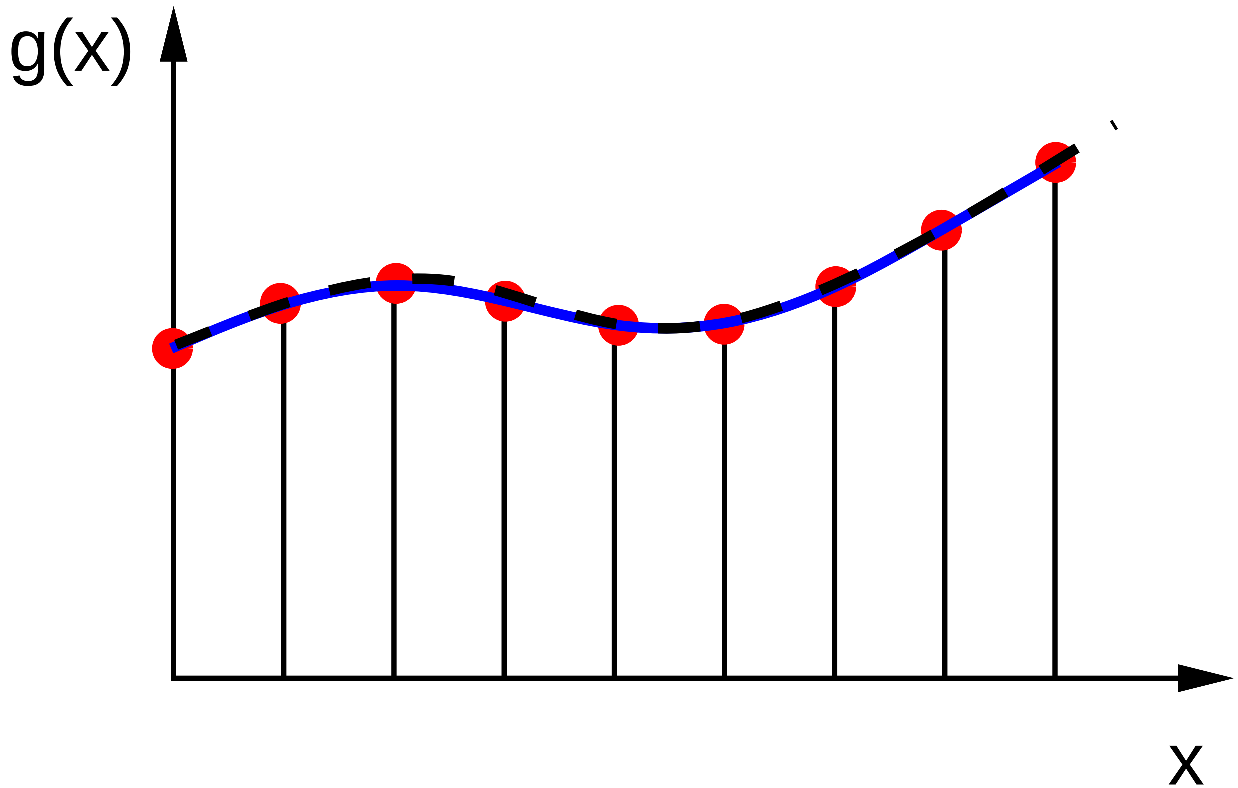
Sampled signal



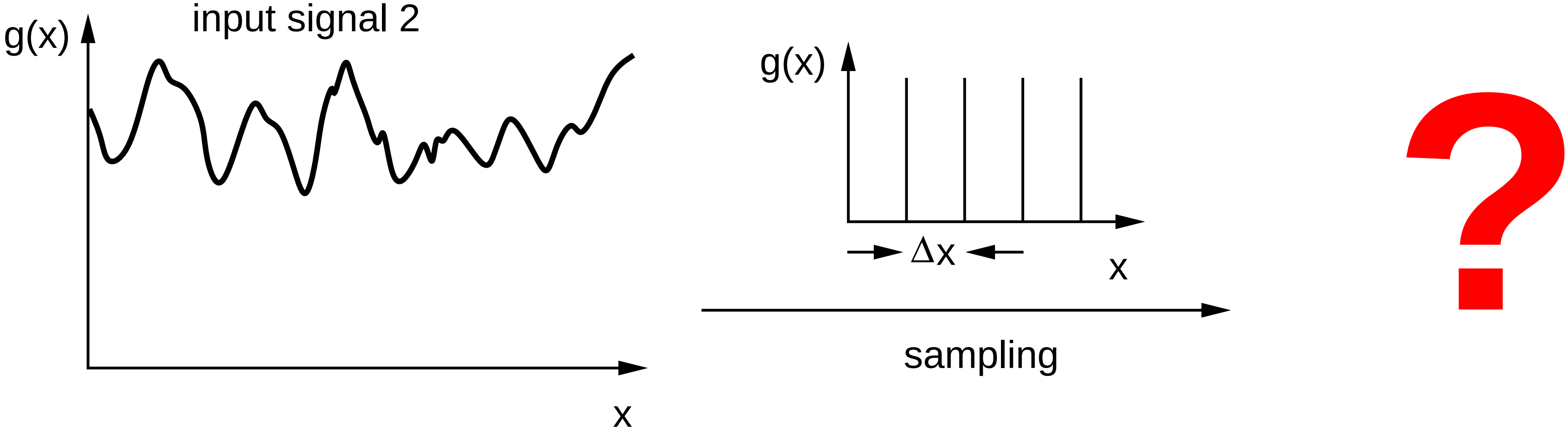
Reconstructed signal (by fitting a smooth function to it)

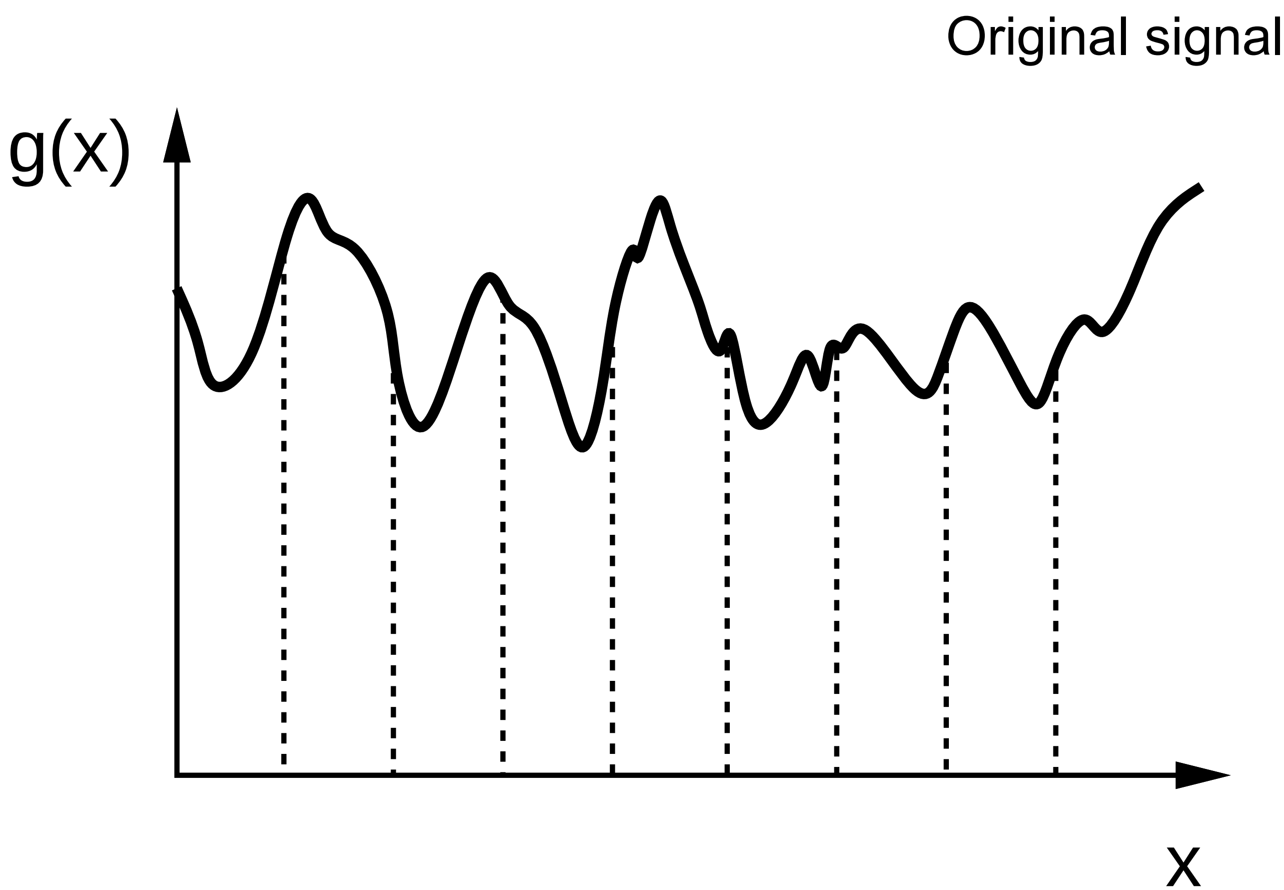


Reconstructed signal superimposed on original signal

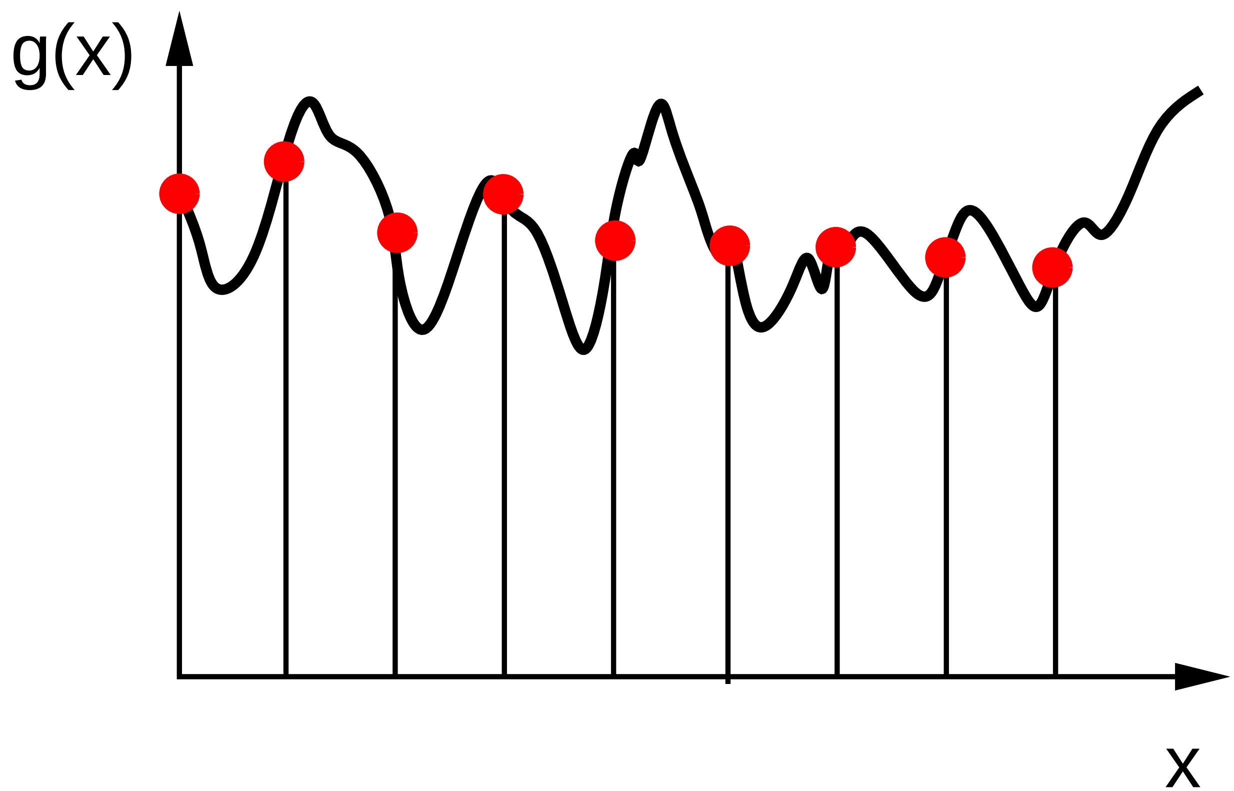


Second, we look at the following signal, which changes much faster:

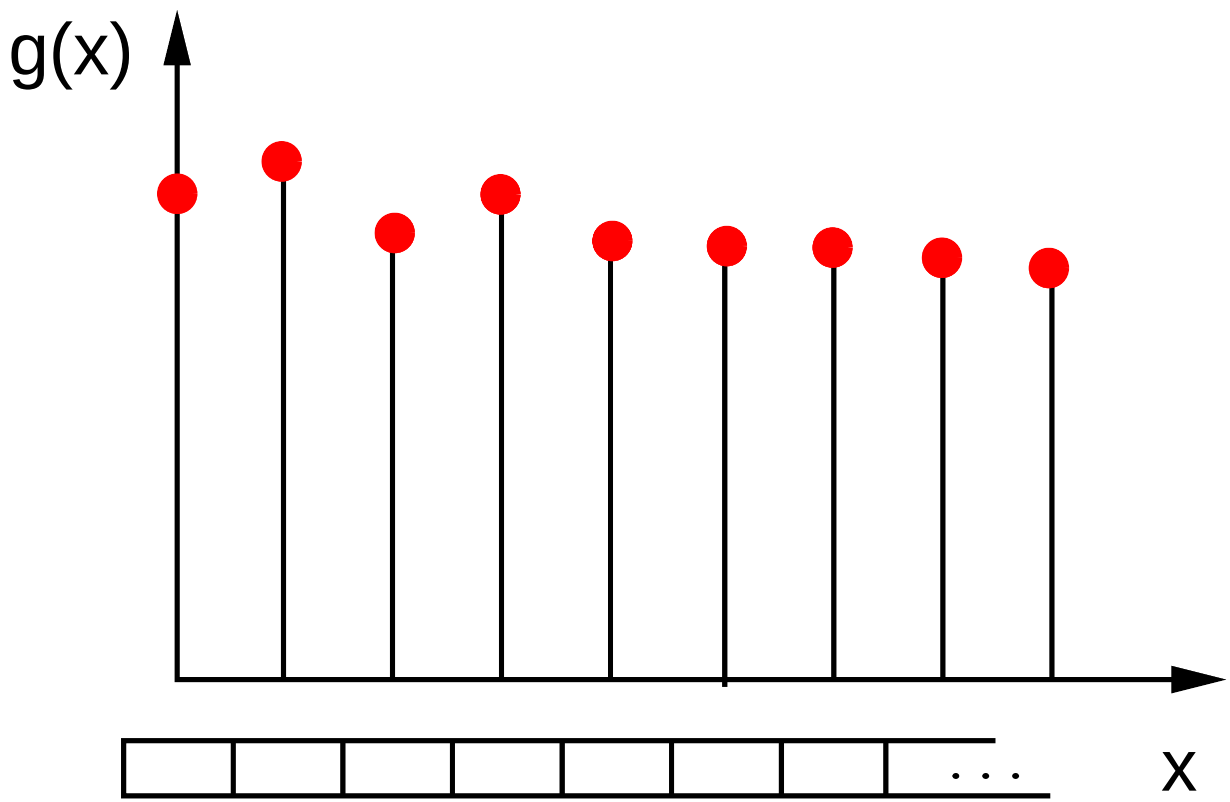



Original signal
g(x)
x

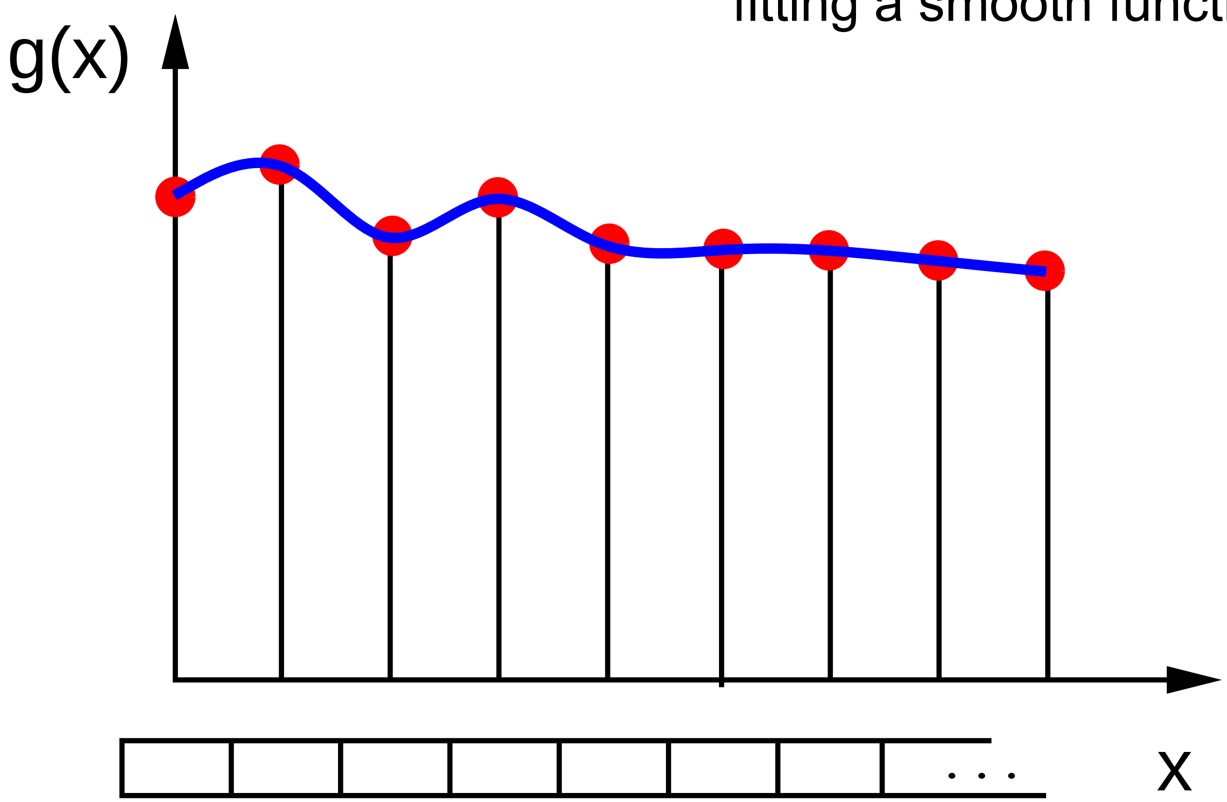
Original signal being sampled



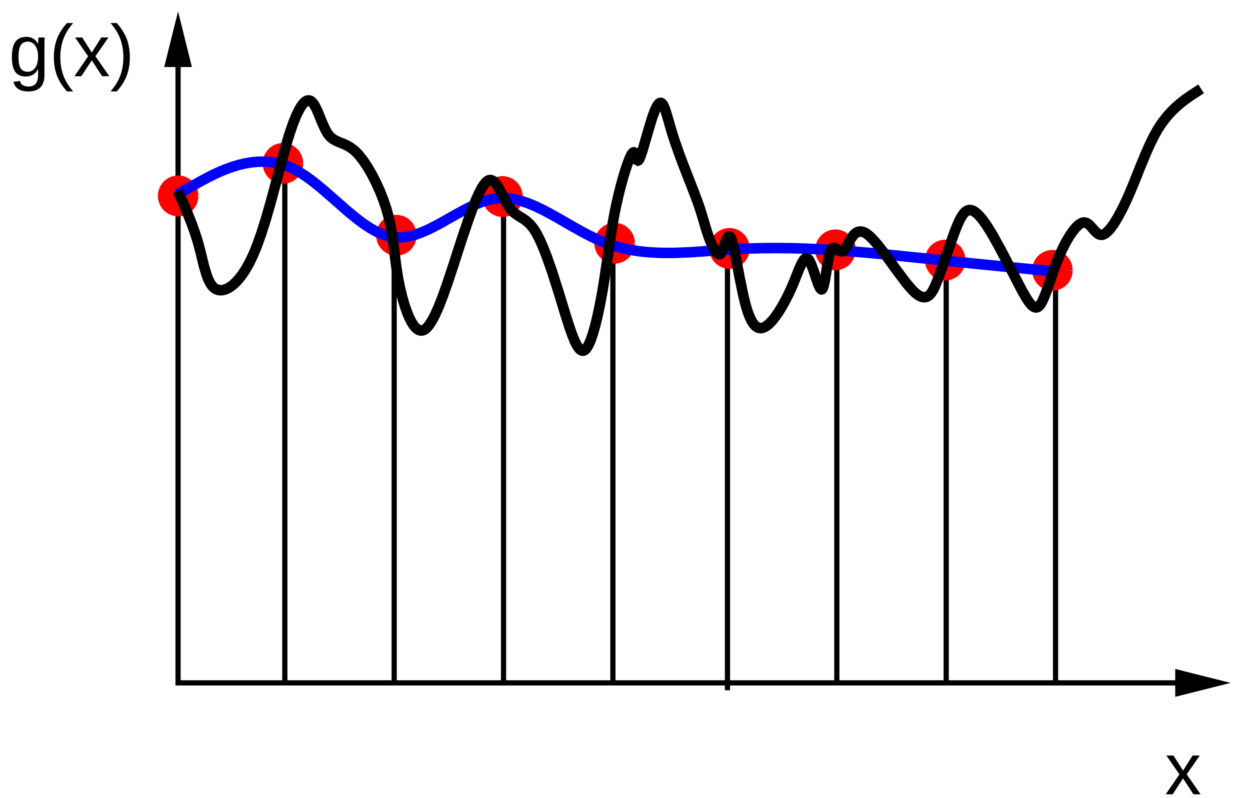
Sampled signal



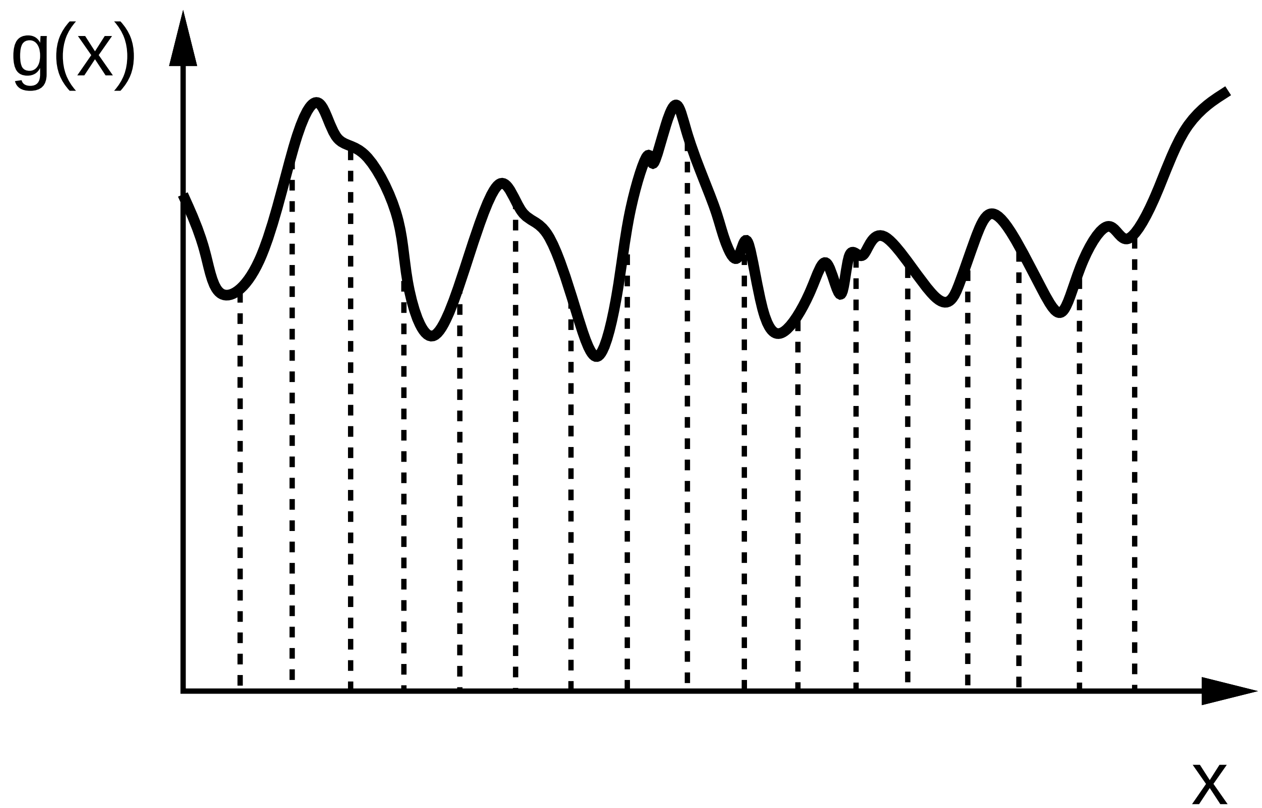
Reconstructed signal (again by fitting a smooth function to it)



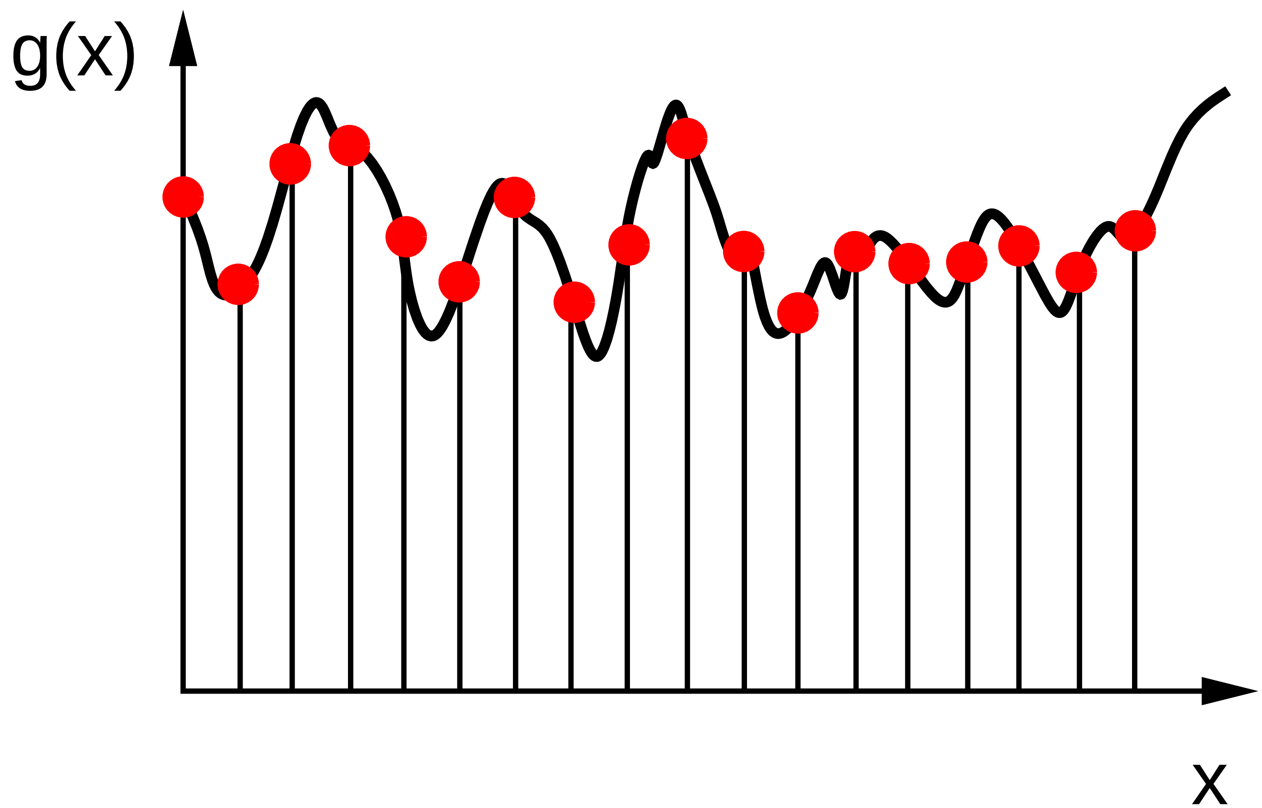
Reconstructed signal superimposed on original signal



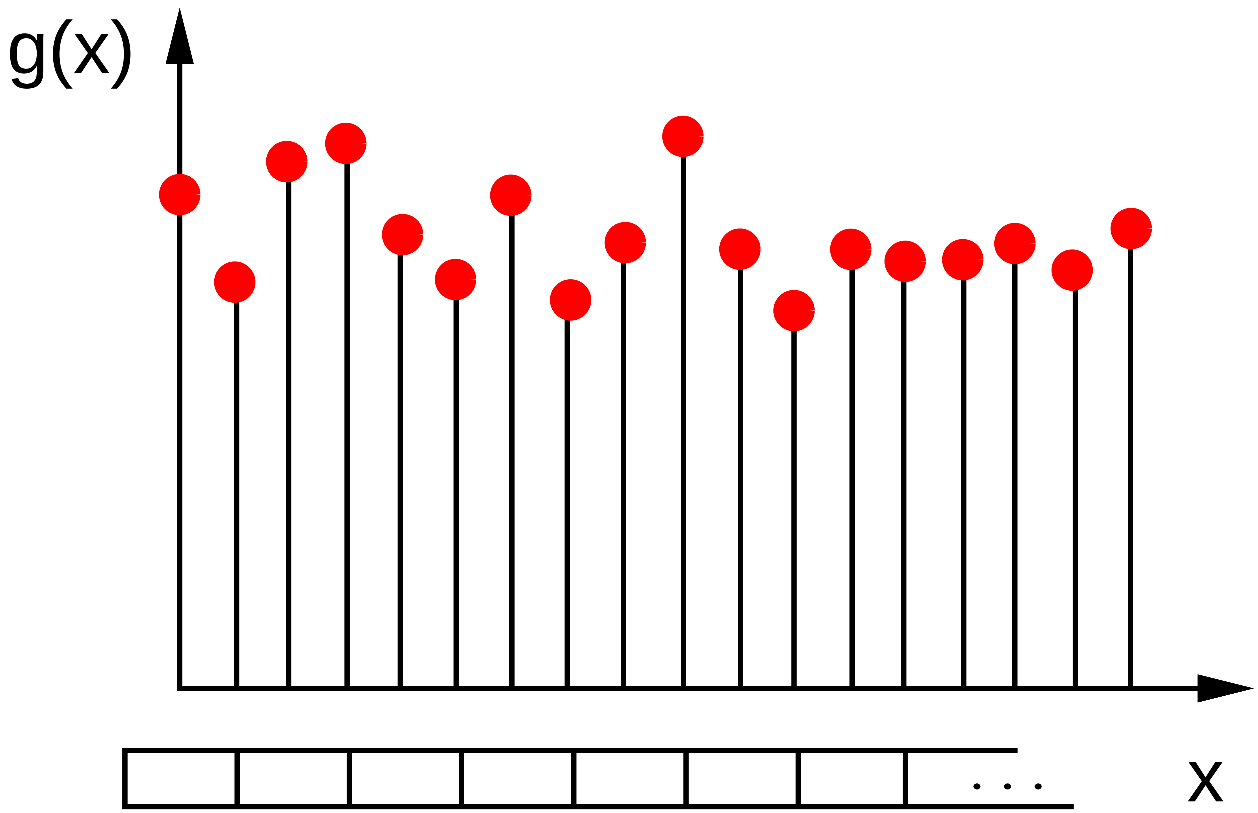
Now, if we reduce the sampling interval, Δx, by half (i.e., doubling the sampling rate) as follows:



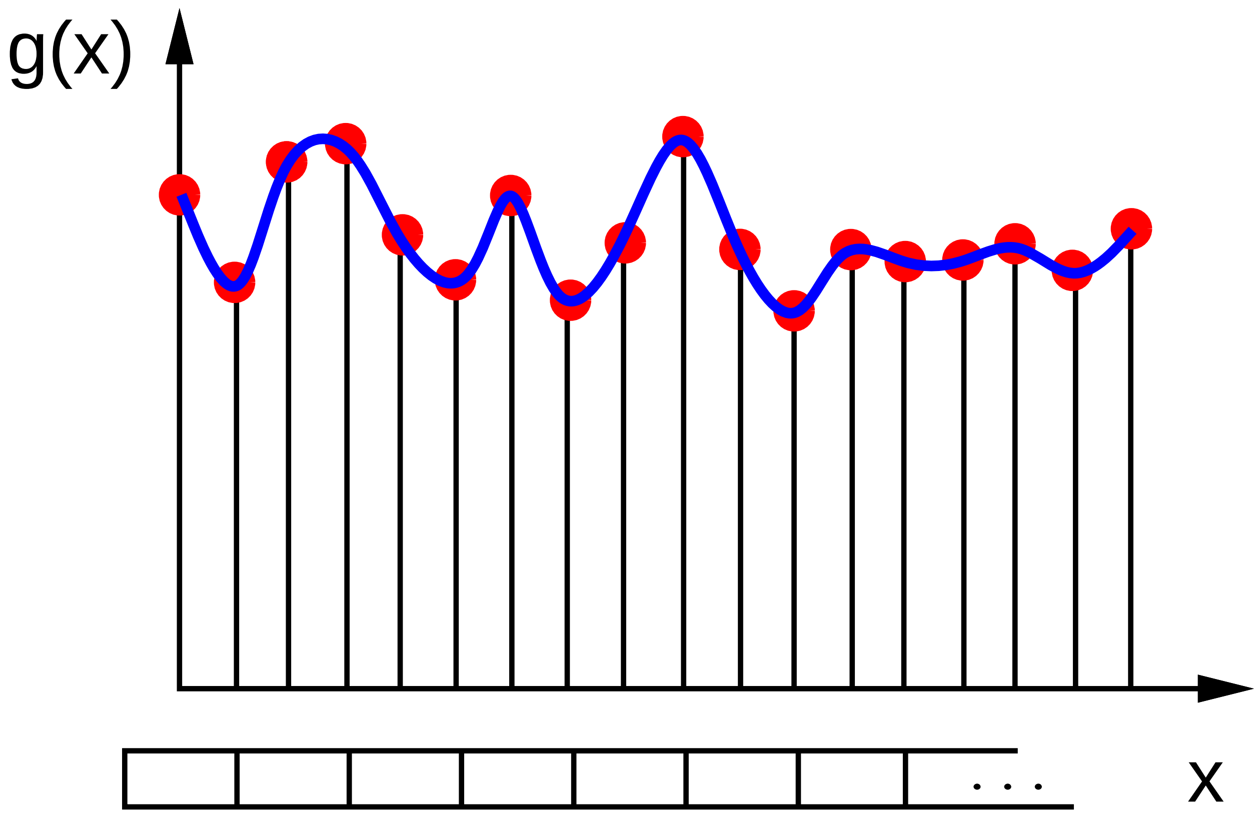
Original signal being sampled



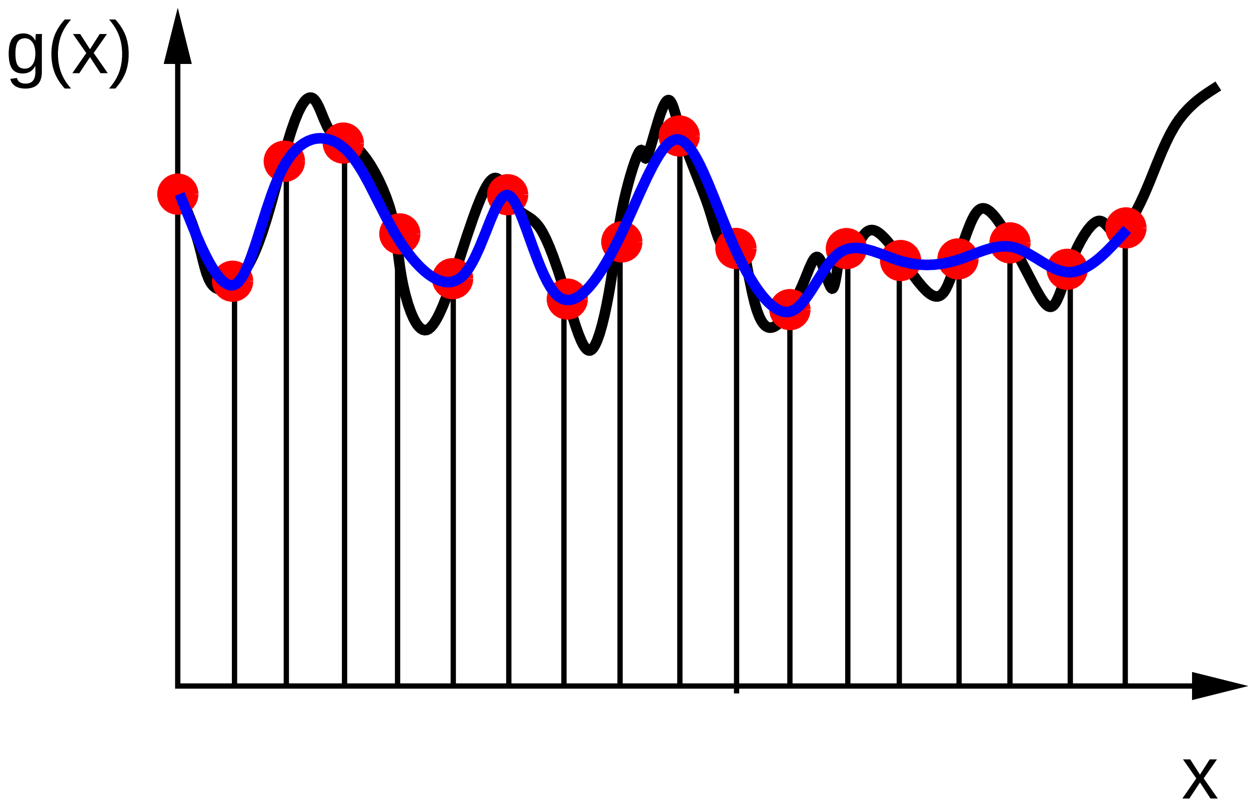
Sampled signal



Reconstructed signal (again by fitting a smooth function to it)



Reconstructed signal superimposed on original signal



- Hence, the sampling interval, *Dx,* should depend on how rapidly the input signal *g*(*x*) is changing against *x*.
- The more rapid the input signal is changing against *x* (i.e., the higher the frequency of the input signal), the smaller the sampling interval (i.e., the higher the sampling frequency) should be used.

- However, as the sampling frequency increases, both the storage space and the processing time will be increased. It may also require more expensive hardware.
- When sending the sampled data through the network, increasing the sampling frequency may also increase the amount of data being sent through the network.
- Hence, given an input signal, we need to find the minimum sampling frequency that would preserve the information of the input signal.

The Sampling Theorem

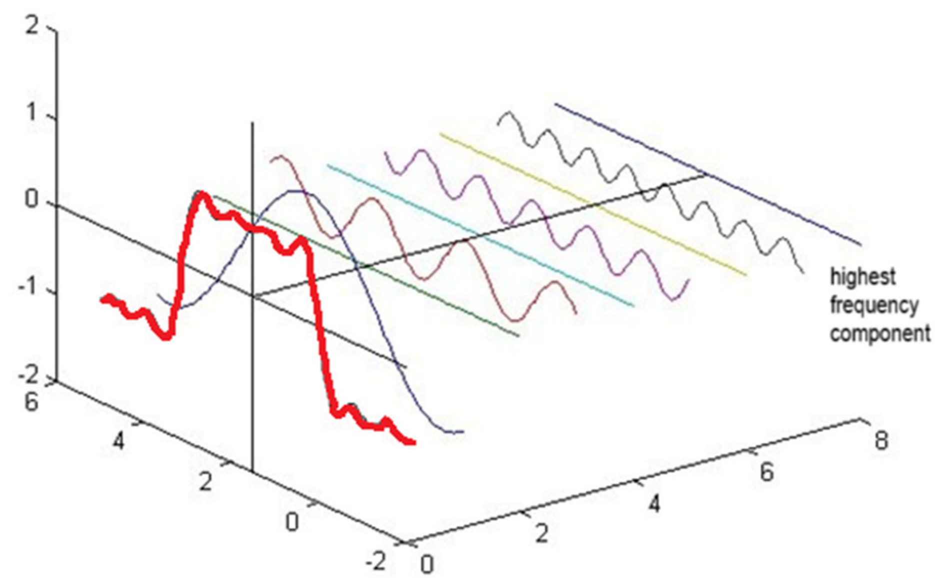
The ***sampling theorem*** states that if the sampling frequency, f_s, is greater than twice of the highest frequency component, f_h, in the input signal, it will be possible to reconstruct the signal from the sampled data without information loss.

Therefore, the sampling theorem states that:

$$f_s > 2 * f_h$$

f_h also defines the bandwidth of the input signal, and it is sometimes referred to as the *Nyquist frequency.* It determines the minimum sampling frequency without information loss.

An example:



A 1D audio signal (e.g., the red curve) can be decomposed into many basic frequency components (four in the above example). The highest frequency component f_h of this audio signal is the right most one.

Data Rates

The file size for a 4-minute song is computed as:

- Human ears can hear sound in the range of 20Hz to 20KHz. (Hz is the unit of frequency.)
- The typical sampling rate for audio CD is set to 44.1KHz, and each sample is 2 Bytes in size.
- The file size for 1 second of the song (stereo) is:

  44,100 * 2 * 2 Bytes = 176,400 Bytes

- The file size for the whole song is then:

  176,400 * 4 * 60 Bytes = 40.4 MB

For images/videos, the sampling frequency depends on the content.

- A lot of changes -> high frequencies
- Smooth changes -> low frequencies.

In practice, it is often difficult to have a sampling rate high enough to satisfy the sampling theorem for videos and images. Hence, it is usually the higher the better. The file size for a 90-minute HDTV movie is:

- Resolution is 1920 x 1080, and each sample takes 3 Bytes.
- The file size for 1 frame (or image) of the movie is:

  1920 * 1080 * 3 Bytes = 5.933 MB

- The file size for the whole movie (25 frames/sec) is then:

  5.933 * 25 * 60 * 90 MBytes = 782 GB

From the above calculations, we can see that:

- A 4K resolution of the above video will take ~3.1TB to store.
- Video files can be difficult to store. They can be even more difficult to transmit through networks.
- Although audio files are much smaller, it may still be difficult if, for example, we want to put a large number of songs into our mobile phone.
- Hence, we need to compress these files, in order to make it easier to store and transmit them.

Filtering

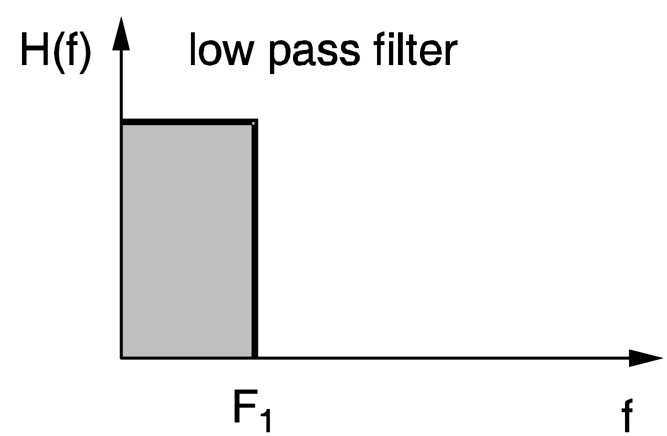
- Given an input analogue signal that we need to digitize, the simplest way is to sample it at a frequency higher than two times the Nyquist frequency (i.e., the highest frequency in the input signal).
- However, this may still generate too much data sometimes.
- Another way is to lower the Nyquist frequency.
- This can be done by removing some high frequencies from the input signal using a *low pass filter*.
  - Bass (low freq.) : https://www.youtube.com/watch?v=FyKcVtQhLVc
  - Soprano (high freq.): https://www.youtube.com/watch?v=6x4k3HISXDk

- A low pass filter is a filter that allows signals below a certain frequency (call it F_1) to pass through. F_1 is referred to as the cut off frequency.

- The ideal low pass filter has the following properties:

$$H(f) = 1 \qquad \text{if } f < F_1$$

$$H(f) = 0 \qquad \text{if } f >= F_1$$



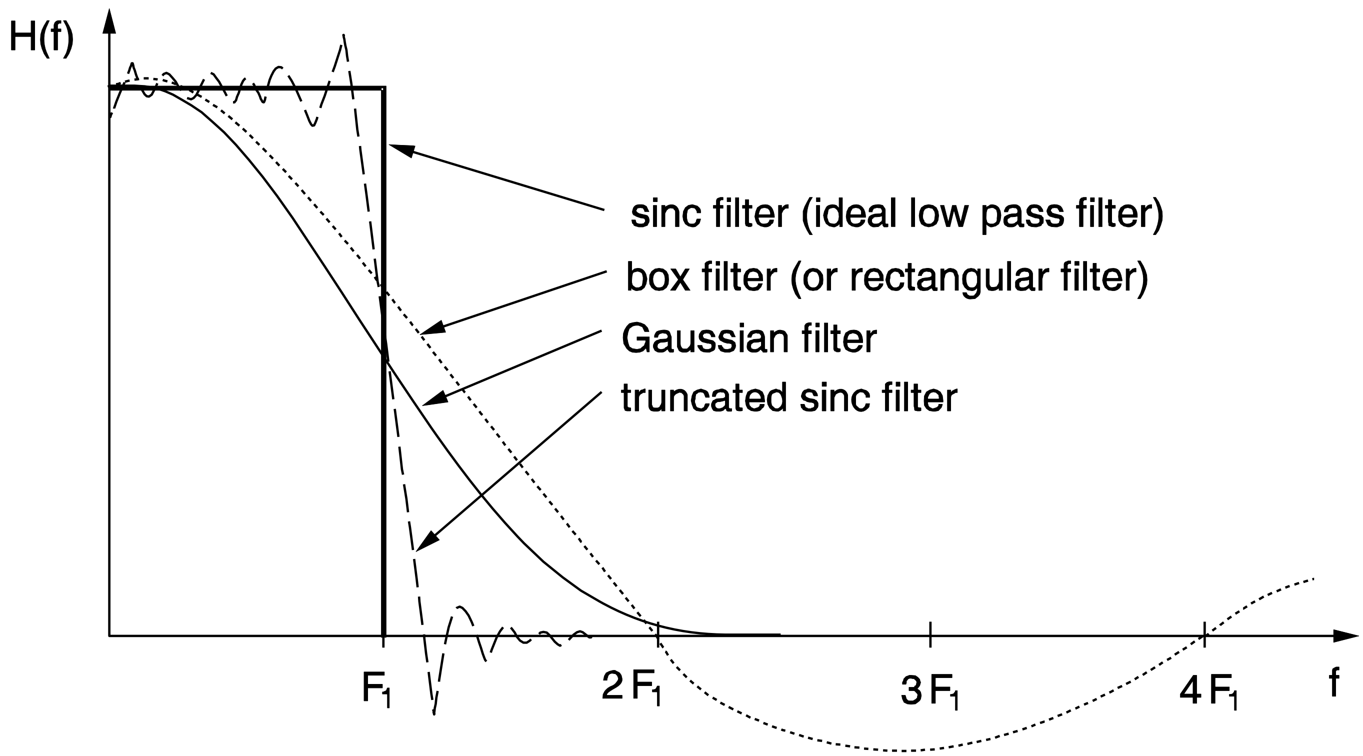
- Then, we have:

$$S_o(f) = S_i(f) * H(f)$$

where $S_i(f)$ is the input signal in frequency domain.
$S_o(f)$ is the output signal in frequency domain.

- As a result, any frequencies in the input signal that are higher than F_1 are removed.
- However, the ideal low pass filter does not exist.

- Practical low pass filters may have the following shapes:



- As we lower the Nyquist frequency, we may also lower the sampling rate.
- As a result, we reduce the amount of output data.
- However, removing the high frequencies is equivalent to throwing away some information that these high frequencies represent, resulting in a loss of information.
- So, the question here is whether these high frequency information are important or not and whether it would affect the audience's perception.

- Telephone service providers use low pass filters to reduce the input signal frequency from the landline/mobile phones to reduce the bandwidth consumption of each phone.
- Their sampling frequency is typically set to 8kHz, meaning that they set F_1 to around 4kHz.
  - A song captured directly from loudspeakers: <audio>
  - The song captured from a phone speaker: <audio>

Network Transmission

- The Internet is known to have low bandwidth in general.
- To transmit video or audio files, we typically compress them first.
- As will be discussed later, most video/audio compression methods compress these files by removing spatial and temporal redundancies from them.
- As a result, the size of the compressed file is not fixed. It depends on the file content. (A video containing a lot of similar frames can be compressed to a smaller size.)

In video streaming, two modes of transmission are available:

- ***Variable bit rate (VBR) transmission***: This mode tries to maintain the quality of the video by letting the data rate fluctuate. The challenges of this mode are in bandwidth estimation and allocation.

- ***Constant bit rate (CBR) transmission***: This mode tries to maintain a constant data rate by adjusting the amount of information removed from the input video. The size of the compressed data can be used as a feedback to control the amount of compression to be achieved by the compressor.

Transmission Errors and Recovery

- To transmit a compressed video through the network, the compressed data are divided into packets and transmitted one by one.
- Interferences and noise may cause errors during the transmission. Some bits of the packets may occasionally be changed by mistakes.
- The bit errors may be detected, and even corrected, by adding parity bits to each packet.

- Assuming that we need to transmit the following package through the network:

⋮
0 1 1 0 1 1 1 0
1 0 0 0 1 0 1 0
0 1 0 0 1 0 0 0
1 1 1 1 0 0 1 1
1 0 0 0 1 0 0 1
⋮

- Two ways of adding parity bits to the package just before transmission: *odd parity* and *even parity*:

⋮	
0 1 1 0 1 1 1 0	0
1 0 0 0 1 0 1 0	0
0 1 0 0 1 0 0 0	1
1 1 1 1 0 0 1 1	1
1 0 0 0 1 0 0 1	0
⋮	

odd parity

⋮	
0 1 1 0 1 1 1 0	1
1 0 0 0 1 0 1 0	1
0 1 0 0 1 0 0 0	0
1 1 1 1 0 0 1 1	0
1 0 0 0 1 0 0 1	1
⋮	

even parity

- The receiving end can detect that there is an error if one of the bits changes its value during transmission:

odd parity

⋮	
0 1 1 0 1 1 1 0	0
1 0 0 0 1 0 1 0	0
0 1 0 0 0 0 0 0	1
1 1 1 1 0 0 1 1	1
1 0 0 0 1 0 0 1	0
⋮	

even parity

⋮	
0 1 1 0 1 1 1 0	1
1 0 0 0 1 0 1 0	1
0 1 0 0 0 0 0 0	0 ← wrong parity value in both cases
1 1 1 1 0 0 1 1	0
1 0 0 0 1 0 0 1	1
⋮	

- If a 2D parity method is used, we may even be able to correct the detected error:

⋮	
0 1 1 0 1 1 1 0	0
1 0 0 0 1 0 1 0	0
0 1 0 0 1 0 0 0	1
1 1 1 1 0 0 1 1	1
1 0 0 0 1 0 0 1	0
0 0 1 0 1 0 0 1	0

odd parity

⋮	
0 1 1 0 1 1 1 0	1
1 0 0 0 1 0 1 0	1
0 1 0 0 1 0 0 0	0
1 1 1 1 0 0 1 1	0
1 0 0 0 1 0 0 1	1
1 1 0 1 0 1 1 0	1

even parity

← parity byte for the data block

What is the limitation with these parity methods?

- If one of the bit changes its value during transmission, we may be able to locate it and then correct it:

⋮	
0 1 1 0 1 1 1 0	0
1 0 0 0 1 0 1 0	0
0 1 0 0 0 0 0 0	1
1 1 1 1 0 0 1 1	1
1 0 0 0 1 0 0 1	0
0 0 1 0 1 0 0 1	0

odd parity

⋮	
0 1 1 0 1 1 1 0	1
1 0 0 0 1 0 1 0	1
0 1 0 0 0 0 0 0	0
1 1 1 1 0 0 1 1	0
1 0 0 0 1 0 0 1	1
1 1 0 1 0 1 1 0	1

even parity

Quality of Service (QoS)

- To provide the user with an acceptable video and audio quality, the ***network*** must provide a certain level of service.
- QoS is a quantitative and qualitative specification on the requirements of an application.
- Packet delay, amount of jitter, bit errors, and packet lost all have significant effects on the quality of the video or audio in a streaming environment.

- Hence, a mechanism is needed for applications to specify their QoS requirements.
- There should be an admission control to determine if a new application should be admitted without affecting the QoS of other on-going applications. It should also make sure that enough resources are available to meet the QoS requirements of the new application and that all the components along the connection agree with and guarantee the QoS requirements.

- Most multimedia applications are dynamic in terms of resource consumption. For example, the data rate (and hence the bandwidth consumption) may change as the user switches the video from one resolution to another.
- There should be traffic policing to make sure that the resouces consumed by each application fall within the agreed QoS specification.
- If an application violates the agreement, another negotiation process on the revised QoS requirements must be initiated.

Wireless Networks

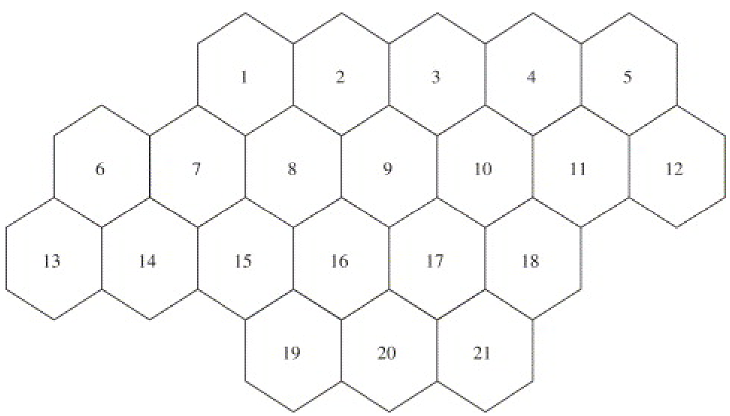
- TV broadcasting in Hong Kong is based on using just 20+ transmitters.

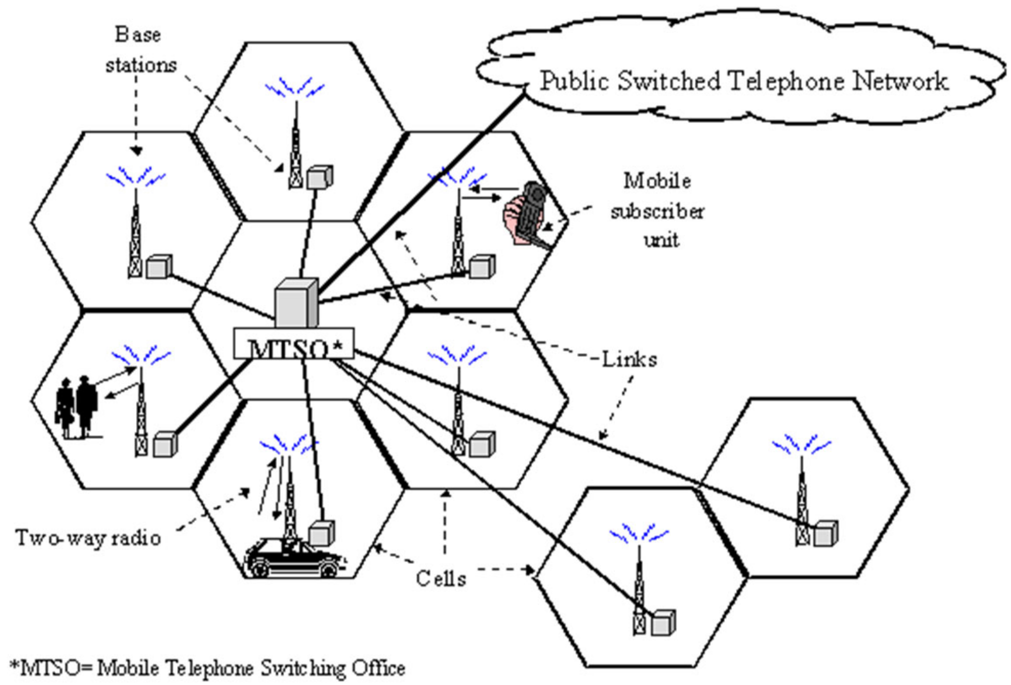
- Each transmitter has high transmission power.

- It is a one-way transmission – stations to homes. Hence, a TV is primarily a receiver.

- Wireless telecommunication services are very different.
- It is a two-way communication. Each phone needs to receive as well as transmit messages.
- With the TV broadcasting approach, each phone will have to use a high power to transmit its messages to the stations, which can be far away. This will use up the battery quickly, and may harm the user.
- Hence, the TV broadcasting approach does not work here. A different approach is needed.

- It is based on dividing Hong Kong into cells. Each cell is very small, and served by a low power transmitter/receiver pair.

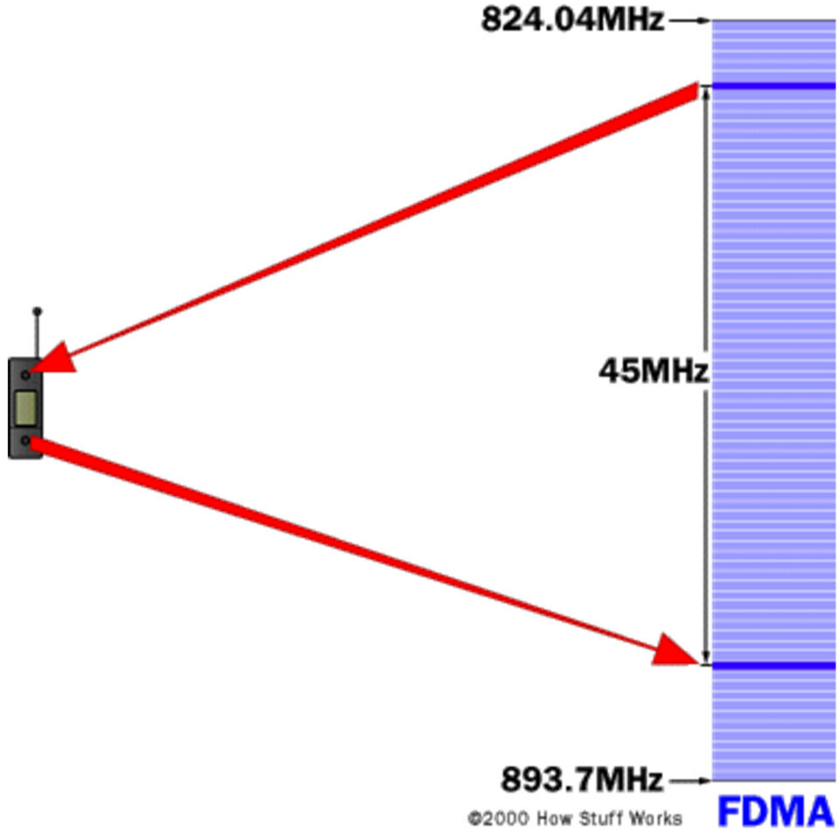




Wireless Networks – 1G

- 1G refers to the first-generation wireless technology.
- It is an analogue telecommunications standard introduced in 1980s.
- Its speed varies between 28kb/s and 56kb/s.
- It is primarily for transmitting voice conversations.

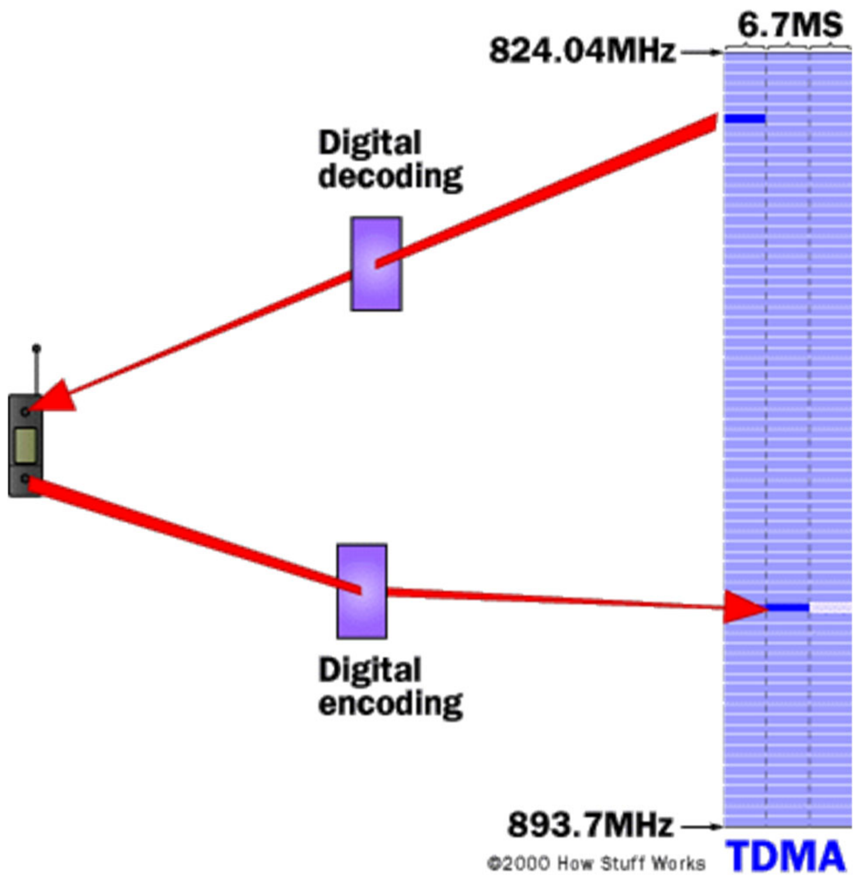
- The input voice signal is modulated to a higher frequency, typically 150MHz or above.
- The modulated signal is transmitted through *Frequency Division Multiple Access* (FDMA), in which each user is assigned a separate frequency channel (of fixed bandwidth) during the call.


824.04MHz
45MHz
893.7MHz
©2000 How Stuff Works
FDMA

Wireless Networks – 2G

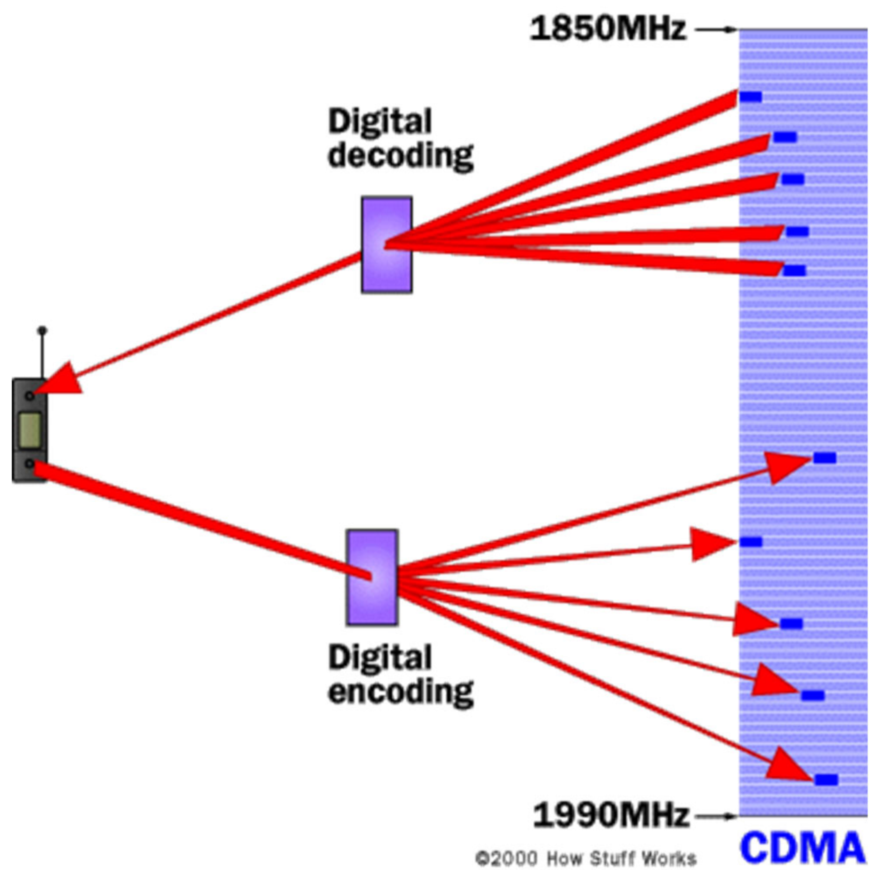
- 2G refers to the second-generation wireless technology.
- It is a digital telecommunications standard introduced in 1991.
- Two main 2G technologies: *Time Division Multiple Access* (TDMA)-based and *Code Division Multiple Access* (CDMA)-based. (It also supports the 1G FDMA.)
- There are a number of 2G standards, including *GSM* (TDMA-based), which accounted for over 80% of all subscribers globally, and *cdmaOne* (CDMA-based), which accounted for about 17% of all subscribers globally.

As the voice data in TDMA is first converted to digital format before transmission, compression can be applied. As a result, the same bandwidth as in FDMA can now be divided into three different time slots, supporting three separate calls.



CDMA divides data from each call into small packets sent through different channels. Multiple calls can now be overlaid on the same channel.

CDMA is slightly more complex than TDMA, but uses the bandwidths more efficiently.



- The main advantages of 2G over 1G include:
  - Phone conversations can be digitally compressed and encrypted.
  - It is significantly more efficient on spectrum usage.
  - It supports data services, such as text messaging, streaming audio and electronic publishing.
- As it is still primarily for voice services, data transmission is rather slow.

Wireless Networks – 3G

- 3G refers to the third-generation wireless technology.
- The standard was introduced in 2001.
- It supports packet-switching, in addition to circuit switching (which is used in traditional telephone systems).
- It provides a data rate of at least 200kb/s. Some even supports mobile broadband access of several Mb/s.
- It supports Web surfing, video mail, mobile multimedia (such as mobile TV and skype), and mobile e-commerce.

- Three common technologies: CDMA2000 (2G CDMA), WCDMA (Wideband CDMA) and TD-SCDMA (Time-division Synchronous CDMA).
- It took much longer for 3G technologies to be adopted globally.
- It is mainly because, unlike 2G networks, 3G networks uses a different radio frequency in order to achieve high data rates. As a result, mobile operators need to build entirely new networks and license new frequencies.

Wireless Networks – 4G

- 4G is based on a set of requirements, together named the *International Mobile Telecommunications Advanced* (IMT-Advanced) specification, introduced in 2013.
- Unlike previous generations, 4G does not support traditional circuit-switched communications. It supports all-IP based communications.
- IMT-Advanced specification specifies a peak data rate of 100Mb/s for high mobility communications (such as from trains and cars) and 1Gb/s for low mobility communications (such as from pedestrians and stationary users).

- Two candidate 4G systems are the *Mobile WiMAX* standard and the *Long Term Evolution* (LTE) standard.
- However, the first-released versions of Mobile WiMAX and LTE support much less than 1Gb/s peak data rates. They are not fully IMT-Advanced compliant.
- Mobile WiMAX Release 2 and LTE Advanced (LTE-A) are IMT-Advanced compliant. They promise data rates of 1Gb/s.

Wireless Networks – 5G

- 5G aims to improve the transmission speed and reduce the transmission latency.
- A key technology is that 5G uses the millimeter-wave spectrum for transmission, which is much faster but can be easily blocked by obstacles.
- Hence, much smaller cells with more focused transceivers are used for communication.
- 5G transmission consumes more energy. To conserve energy, 5G phones may use 5G technologies when needed and 4G technologies otherwise.

Multimedia Streaming

- In traditional file transfer, the whole media file is first transmitted to the client before playing.
- This is simple but has a potentially long delay. Browsing can be difficult.
- Streaming is the transmission of a media file (audio, video or both) from a server to a client, but allows the client to start playing before the file has been completely received.

- Characteristics of multimedia streaming:
  - Typically sensitive to delay

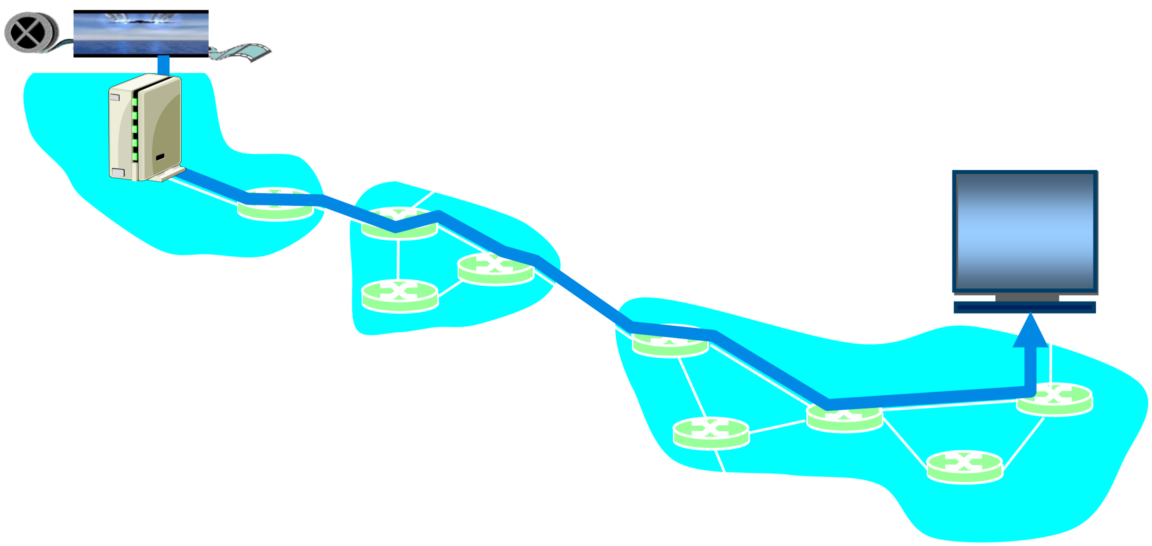
    e.g., difficult to talk to each other on skype under delay.
  - But tolerate occasional packet lose

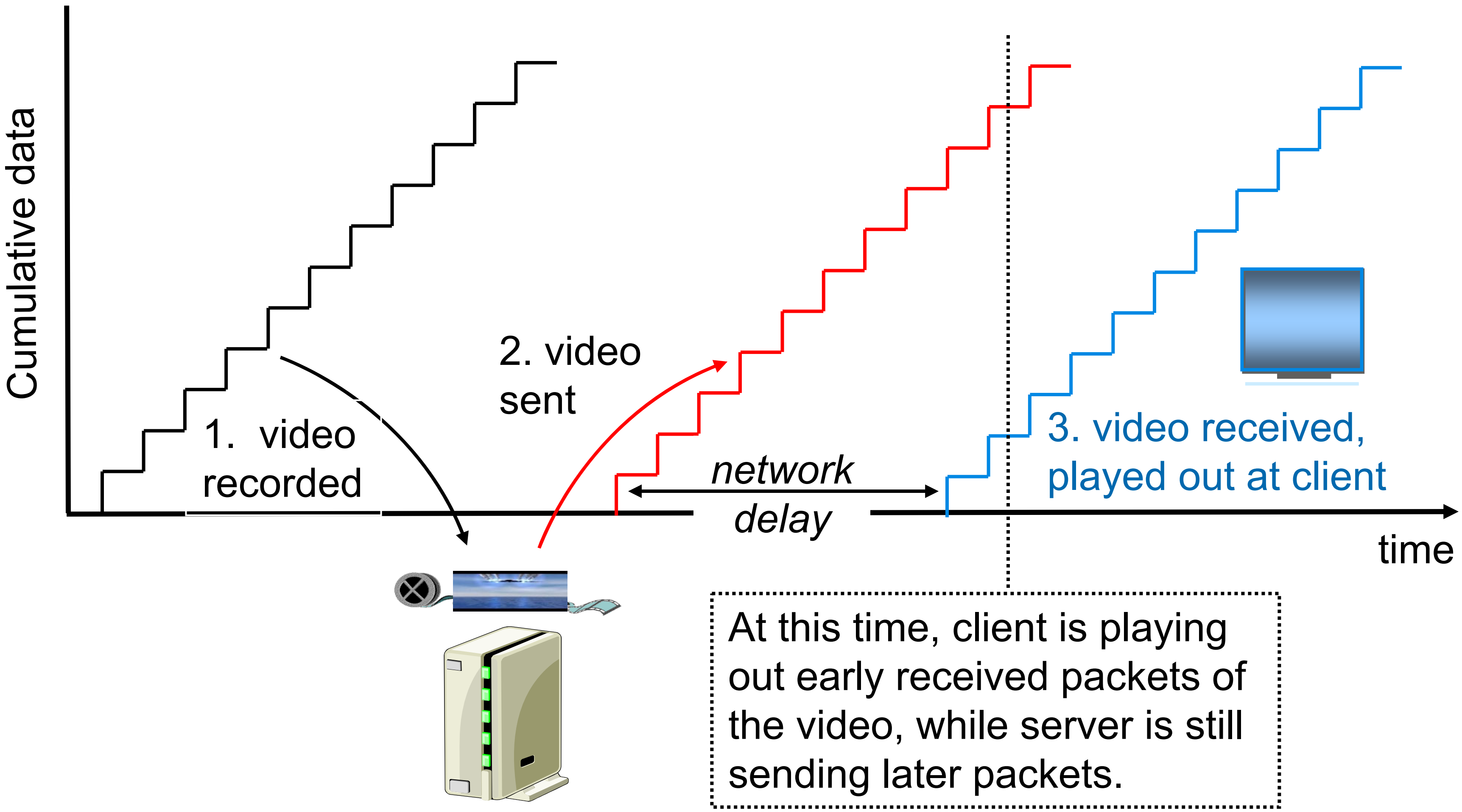
    e.g., OK to most users for occasionally losing a few frames.
  - This is in contrast to traditional data transfer, where packet lose cannot be tolerated but delay can be tolerated.

- Three types of streaming applications:
  - Streaming of stored media
  - Streaming of live media
  - Real-time interactive media

Streaming of Stored Media

- Applications include video-on-demand (VOD).
- A client requests for a video from a video server.
- The video stored in the video server is then streamed to the client.

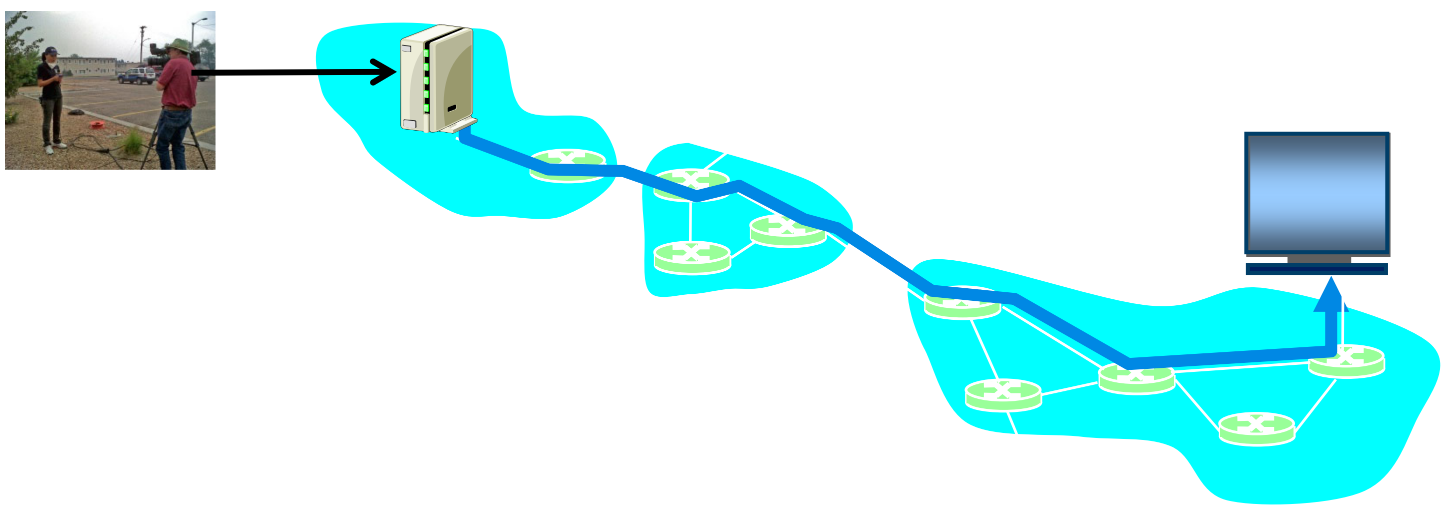



Cumulative data
1. video recorded
2. video sent
network delay
3. video received, played out at client
time
At this time, client is playing out early received packets of the video, while server is still sending later packets.

- The media player at the client may support typical VCR functions, such as Play, Pause, Fast Forward, and Rewind.
- The **RTSP** (Real-Time Streaming Protocol) is typically used in writing the streaming algorithm.
- Limitation: Packets need to have arrived at the client before they can be played. Hence, Fast Forward may cause problems – not enough data in the local buffer.

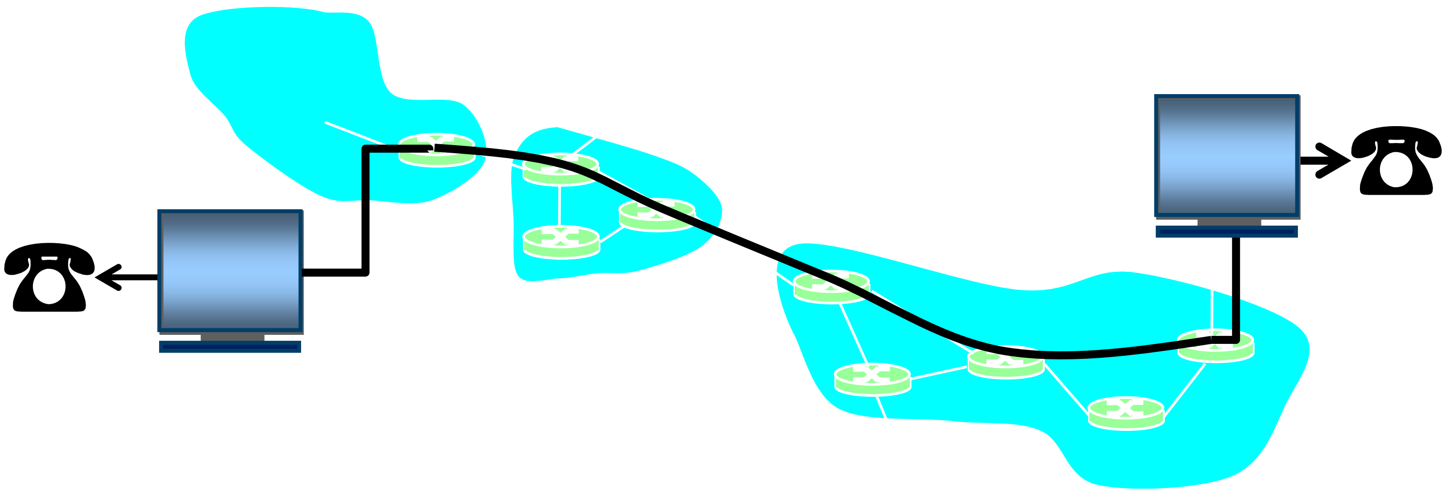
Streaming of Live Media

- Applications include Web TV and Web Radio services.
- A typical time lag of a few seconds for server compression, transmission, client buffering and client decompression.
- Can support Play, Pause, Rewind, but not Fast Forward.



Real-Time Interactive Media

- Applications include IP phones, skype, and Zoom.
- Video and audio captured at a client are compressed and then transmitted immediately.



- To provide a better user experience, ***end-to-end delay*** should be smaller than 150ms up to 400ms for audio, including compression, transmission, and decompression.
- To reduce the time delay, compression/decompression algorithms cannot be too complex – resulting in lower video quality.
- A small buffer can be used at each client to smooth the output. However, buffer causes more delay – a tradeoff.

TCP/UDP

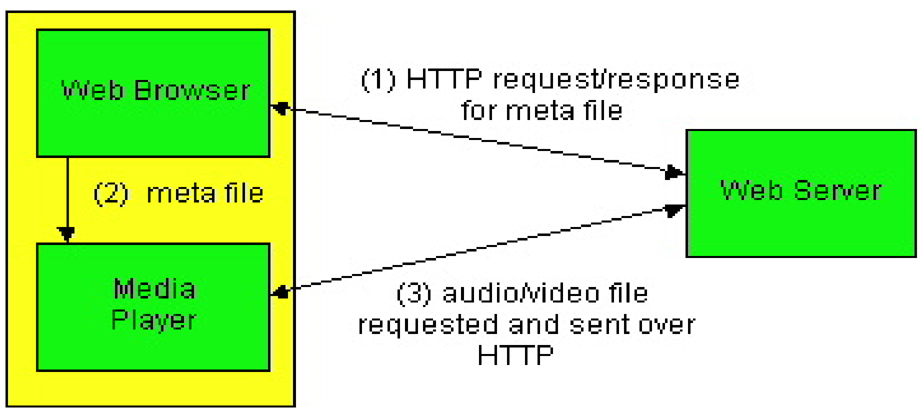
- The Transmission Control Protocol (TCP) is used to establish a reliable connection between the client and the server in order to support file transfer.
- TCP guarantees that the client will receive all packets sent by the server. If a packet is lost, the server will automatically resend it.
- It also guarantees that the packets will be received in the order that the server sends them, by numbering all packets sent.
- As a result, TCP has high overheads and not efficient.

- The User Datagram Protocol (UDP) sends each packet as a datagram to the client.
- Each datagram is similar to a TCP packet but contains full sender/receiver address information. So, datagrams can be sent via different routes to the client.
- UDP does not include error-checking.
- Hence, UDP is more efficient.

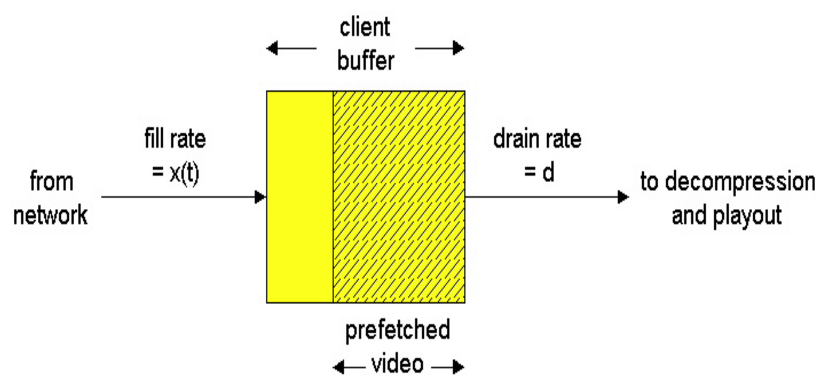
	TCP	UDP
Transmission Method	Establish a connection first before sending packets	No connection is set up. Datagrams contain full address information, and each may be sent via different routes.
Reliability	Ensure that all packets will be received by the client, unless the connection is lost.	No guarantee that all datagrams will be received nor will they be error-free.
Ordering	Guarantee that the first packet sent will be received before the second.	The datagrams can be received in any order, disregard of their sending order.
Overheads	High overheads to ensure ordering and error-free.	Low overheads, due to simplicity.
Applications	HTTP is usually based on TCP. Multimedia streaming may occasionally use TCP.	Multimedia streaming is usually based on UDP. HTTP may occasionally use UDP.

Streaming Basic

- Streaming can be done over HTTP as:
  - Browser GETs metafile
  - Browser launches media player and passes metafile to the player
  - Media player contacts server
  - Server streams the requested media to the media player

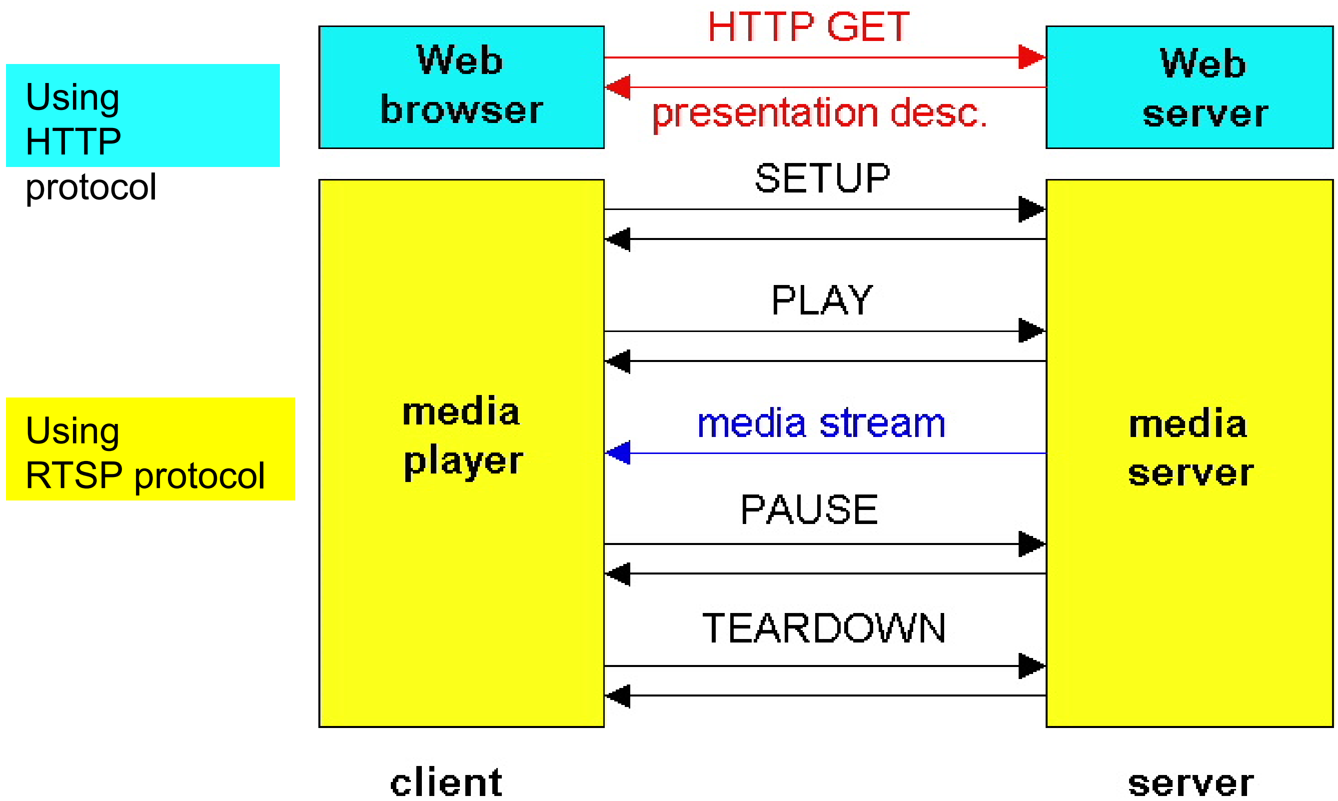


- Client-side buffering is to compensate for the network delay jitter. A higher network delay fluctuation requires a larger buffer, resulting in a higher end-to-end delay.



- Server may store multiple copies of each video, compressed at different rates, for clients with different network bandwidths.

The Real Time Streaming Protocol (RTSP)



- Web browser GETs metafile from Web server.
- Web browser launches media player.
- Media player sets up RTSP control connection and data connection to media server.

- HTTP is not targeted for multimedia content.
- HTTP does not support functions such as fast forward.
- RTSP supports Play, Pause, Record, Repositioning, etc.
- RTSP supports media to be transported over TCP or UDP.
- RTSP adopts out-of-band control. RTSP control messages use a different port number than the media stream. While the media stream is considered as in-band, the control messages are considered as out-of-band.

An Example – IP Phone

- A speaker usually alternates between talk spurts and silent period.
- During talk spurts:
  - 64Kbps (or 8KBps) data rate
  - Sends one chunk of data every 20ms, i.e., 160 Bytes data.
  - A chunk plus header encapsulated into a UDP segment for transmission.
- No need to send UDP segments during silent periods.

- Two kinds of datagram losses:
  - ***Network loss***: lost of datagrams due to network congestion (e.g., router buffer overflow).
  - ***Delay loss***: datagrams arrive too late to be played by the media player, due to higher-than-expected end-to-end delay. (Typical maximum allowable end-to-end delay, D_{max}, is ~400ms.)
- A larger D_{max} results in smaller delay losses, but less interactive experience.
- Usually, up to 10% of datagram losses can be tolerated.

Recovery of Datagram Loss: Scheme 1

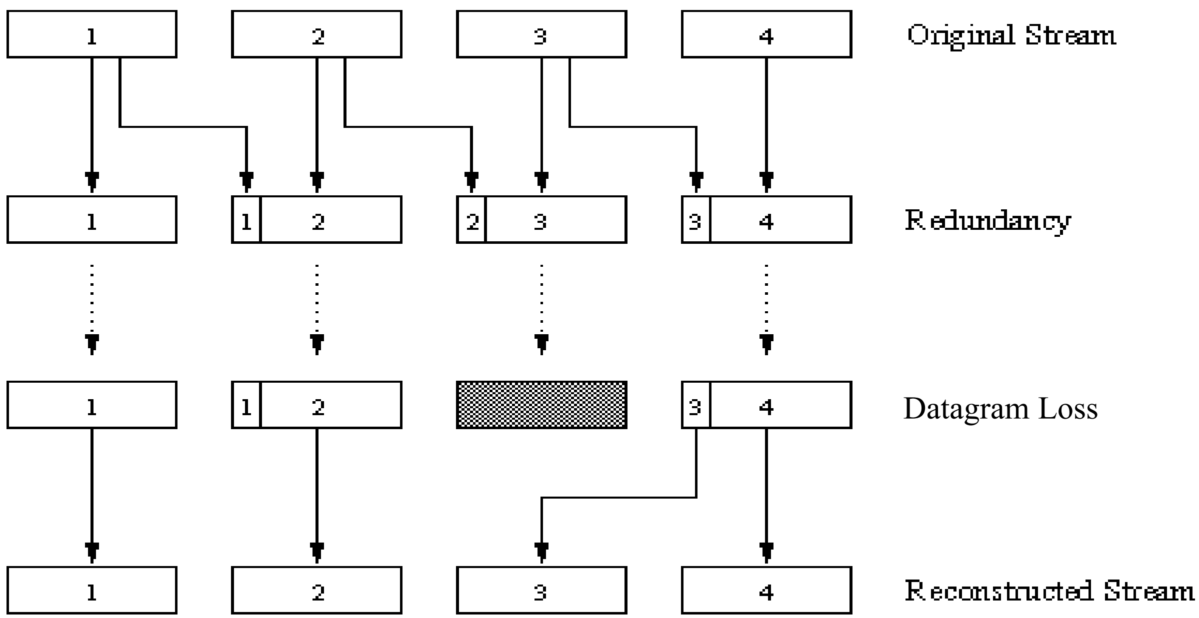
- We may use a method similar to the parity method here.
- Given a group of *n* chunks, we may use the same bits from different chunks to create a parity bit. This will produce an additional chunk of parity bits.
- Now, we need to send out a total of *n+1* chunks, increasing the bandwidth requirement by *1/n*.
- If one of the *n* chunks of media data is lost, we can reconstruct the lost chuck from the received *n* chunks.

What would happen if the lost chuck is the parity chunk?

- The tradeoff for the size of *n*:
  - A smaller *n* → more bandwidth is wasted due to the parity chunk.
  - A larger *n* → a higher probability that two or more chunks will be lost.

Recovery of Datagram Loss: Scheme 2

- We generate a low quality version of each chunk, e.g., 13kbps versus 64kbps for the original.
- This low quality version is packed to the following chunk.



Recovery of Datagram Loss: Scheme 3

- Chunks are divided into smaller units.
- Each datagram contains units from multiple chunks.
- If a datagram is lost, we still have most of every chunk.
- Advantage: no redundancy, but higher delay.