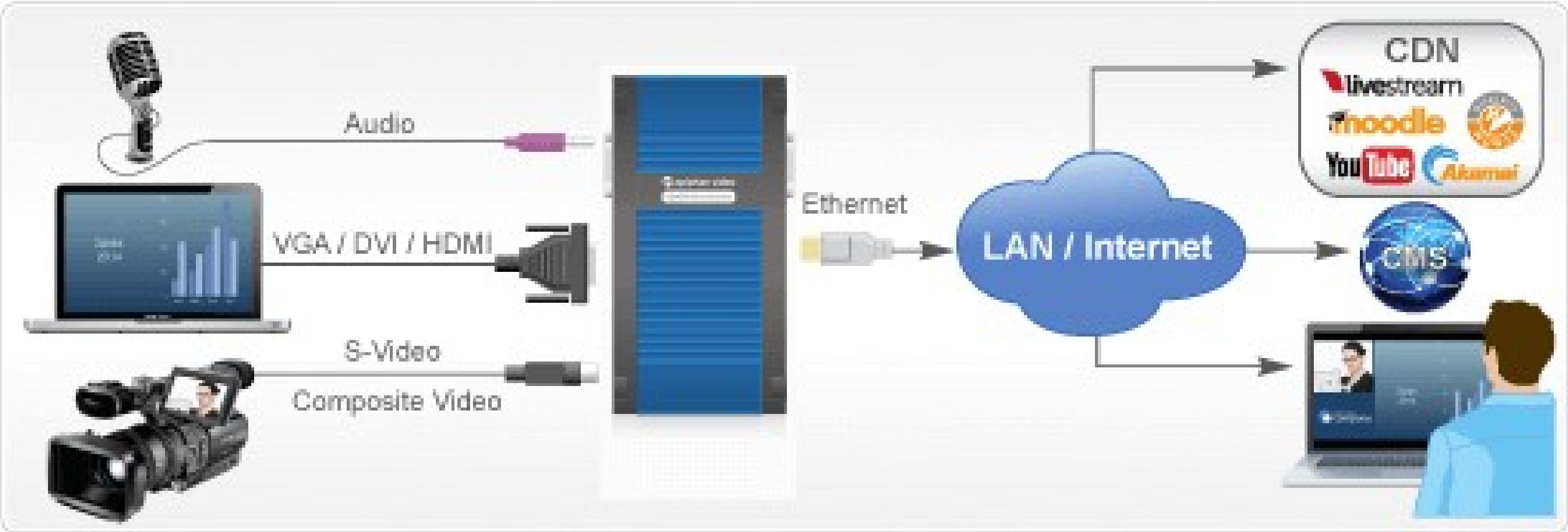
Multimedia Networking

School of Informatics
Institute for Computing Systems Architecture

2. Streaming Live Audio and Video (Introduction)

- Live streaming comes from a content source such as video cameras and microphones. It is made available at the same time as the event being filmed occurs.
- Similar to traditional broadcast radio and TV, except that transmission is on Internet.



2. Streaming Live Audio and Video (Features)

- Applications
  - Internet Radio (Talk Show)
  - TV broadcast (News, TV shows)
  - Live sporting event, IPTV (pplive, ppstream etc)
- Streaming (as with streaming *stored* multimedia)
  - Playback buffer
  - Playback can lag tens of seconds after transmission
  - Delay constraints more stringent than streaming stored video but less than conversational voice/video
- Interactivity
  - Fast forward (impossible) BUT Rewind, pause (possible)
- Continuity
  - Average throughput greater than video rate desired
  - Forward error correction (FEC) more effective than reactive loss recovery

School of Informatics
Institute for Computing Systems Architecture

2. Streaming Live Audio and Video (Distribution)

- Distribution of live audio/video to many receivers can be done via:
  - Network layer approaches
    - Multiple unicast streams
    - IP multicast streams
  - Application layer approach
    - Multicast using P2P networks or CDNs.
  - **Network layer approaches**
  - Multiple Unicast streams
  - One-to-one connection between the server and a client
  - Each client receives a distinct stream
  - Burden upon the system and links (due to redundant packet generation and transmission) but enable interactivity.

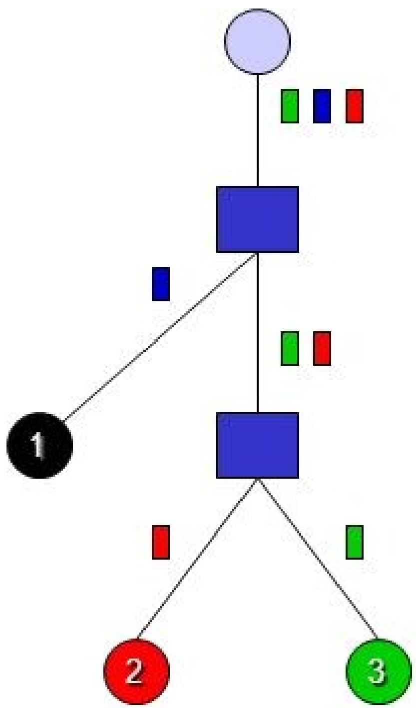
School of Informatics
Institute for Computing Systems Architecture

2. Streaming Live Audio and Video (Distribution)

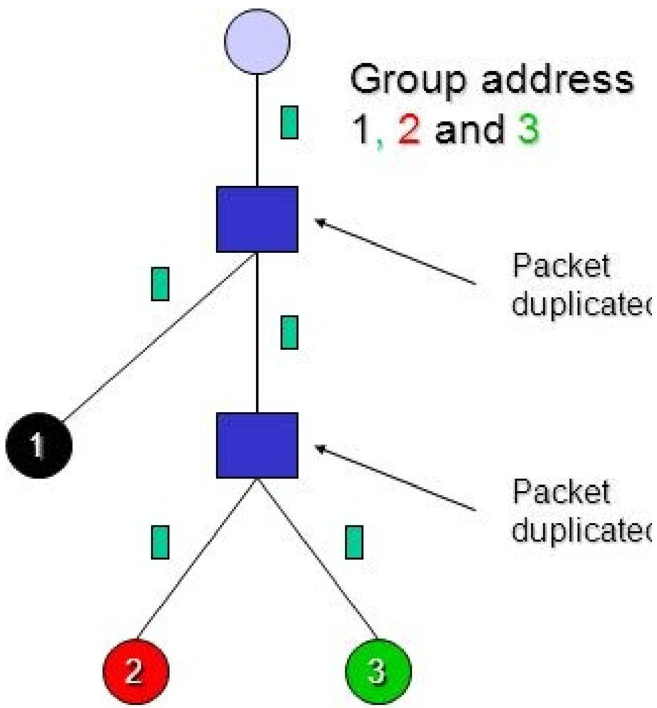
- IP multicast streams
  - A single flow, from server to a group of clients
  - IP packets have multicast address in class D
  - Minor burden on server (sends a single copy)
  - Better capacity utilization of network (single copy of message on a link)
  - Requirements
    - Group management
    - Packet replication at network nodes (from a single input port to many output ports).
  - All clients receive the same stream and do not have control of content playback

2. Streaming Live Audio and Video (Distribution)

Multiple Unicast:



Multicast:

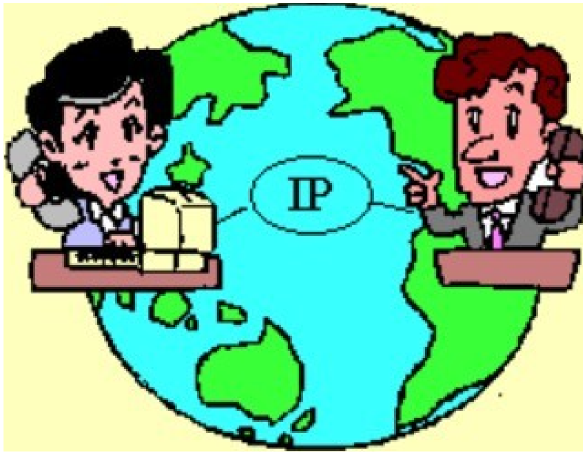


3. Conversational Voice and Video over Internet

- Includes group meetings involving more than two participants
- Voice/video generated by every member (E.g., Skype, Google Talk)

3. Conversational Voice and Video over Internet

- We will focus on **conversational voice over Internet / Internet telephony / Voice-over-IP (VoIP)**
  - Real-time conversational voice over the Internet (i.e. Internet Telephony)
  - Protocols: RTP, SIP



3. Conversational Voice and Video over Internet (VoIP: Characteristics)

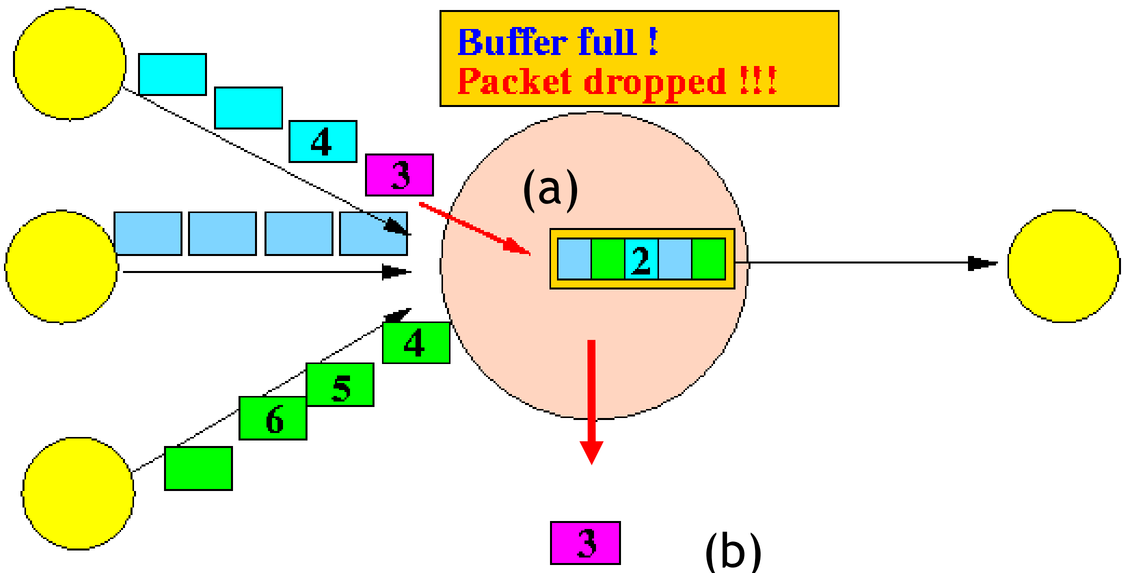
- Digitised voice encapsulated in packets and transported between 2 or more VoIP call participants
- *Highly delay sensitive*
  - ≤ 150ms ideal
  - > 400ms unacceptable
- *Loss-Tolerant*
  - Occasional glitches in vedio/audio playback
  - Loss recovery schemes and error concealment: FEC, interleaving

3. Conversational Voice and Video over Internet (VoIP: Best Effort delivery and its limitations)

- IP provides Best-Effort delivery:
  - Service tries best to move a packet from source to destination
  - BUT does not guarantee speedy delivery, no-packet loss
- Limitations of best-effort IP service in the context of VoIP
  - Example:
    - The sender generates bytes at a rate of 8,000 bytes per second; every 20 msecs the sender gathers these bytes into a chunk.
    - A chunk with header are encapsulated in a UDP segment. The number of bytes in a chunk is (20 msecs)(8,000 B/sec) = 160B.
    - If constant end-to-end delay, then packets arrive at the receiver periodically every 20 msecs.
    - If unfortunately delay is varying and network is congested, then some packets will arrive with delay and some packets will be lost

3. Conversational Voice and Video over Internet
(VoIP: Best Effort delivery and its limitations)

- **Packet Loss**
  - TCP can solve it by retransmission BUT not unacceptable for conversational VoIP application. Most existing VoIP applications use UDP.
  - Also loss from packet 1 – 20% can be tolerated depending how loss encoded, transmitted and concealed (FEC)
  - If loss is larger than what is tolerable, nothing can be done to recover from bad quality

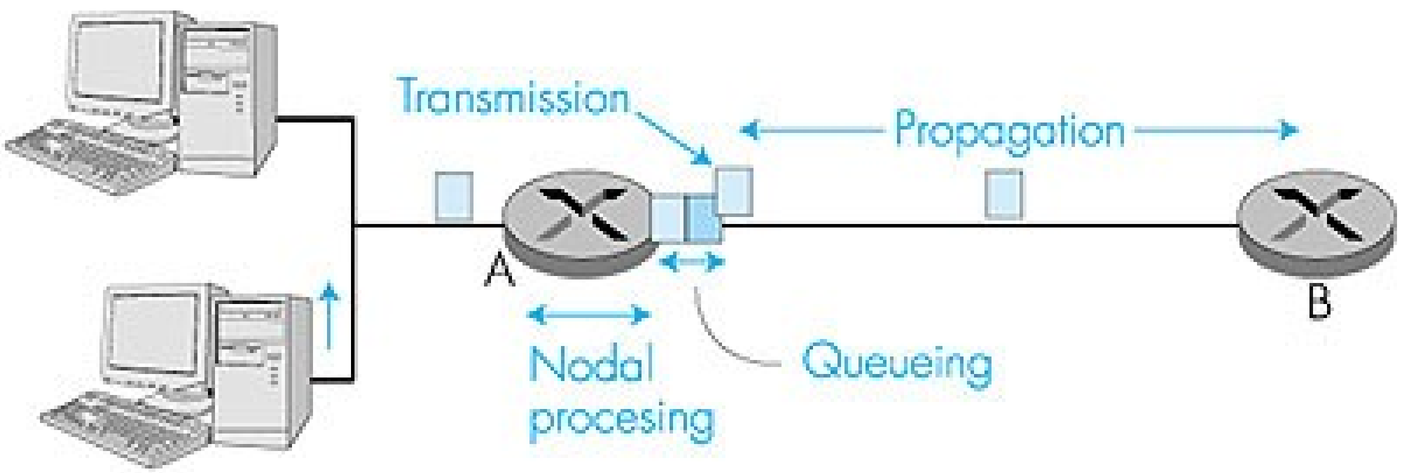


School of Informatics
Institute for Computing Systems Architecture

3. Conversational Voice and Video over Internet (VoIP: Best Effort delivery and its limitations)

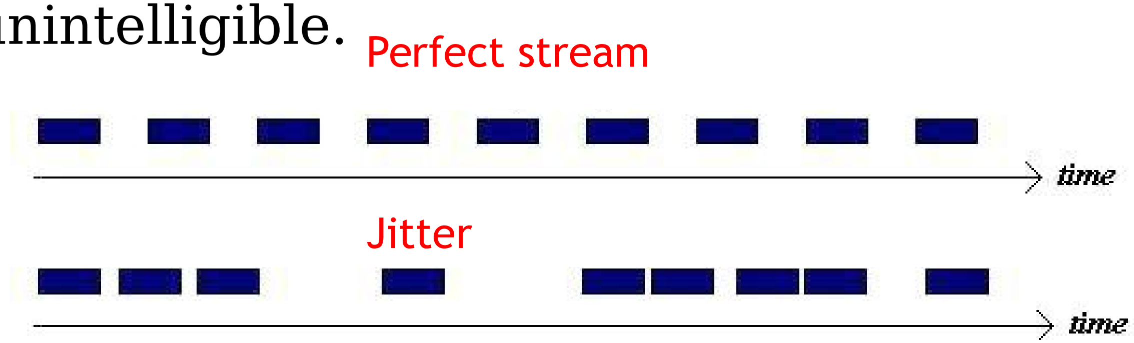
– **End-to-End Delay**

- End-to-end delay is the sum of:
    - Transmission, processing and queuing delay at each router
    - Propagation delay of each link on the path
    - End-system processing delays
    - Packets with delay > 400msec are disregarded by receiver.



3. Conversational Voice and Video over Internet (VoIP: Best Effort delivery and its limitations)

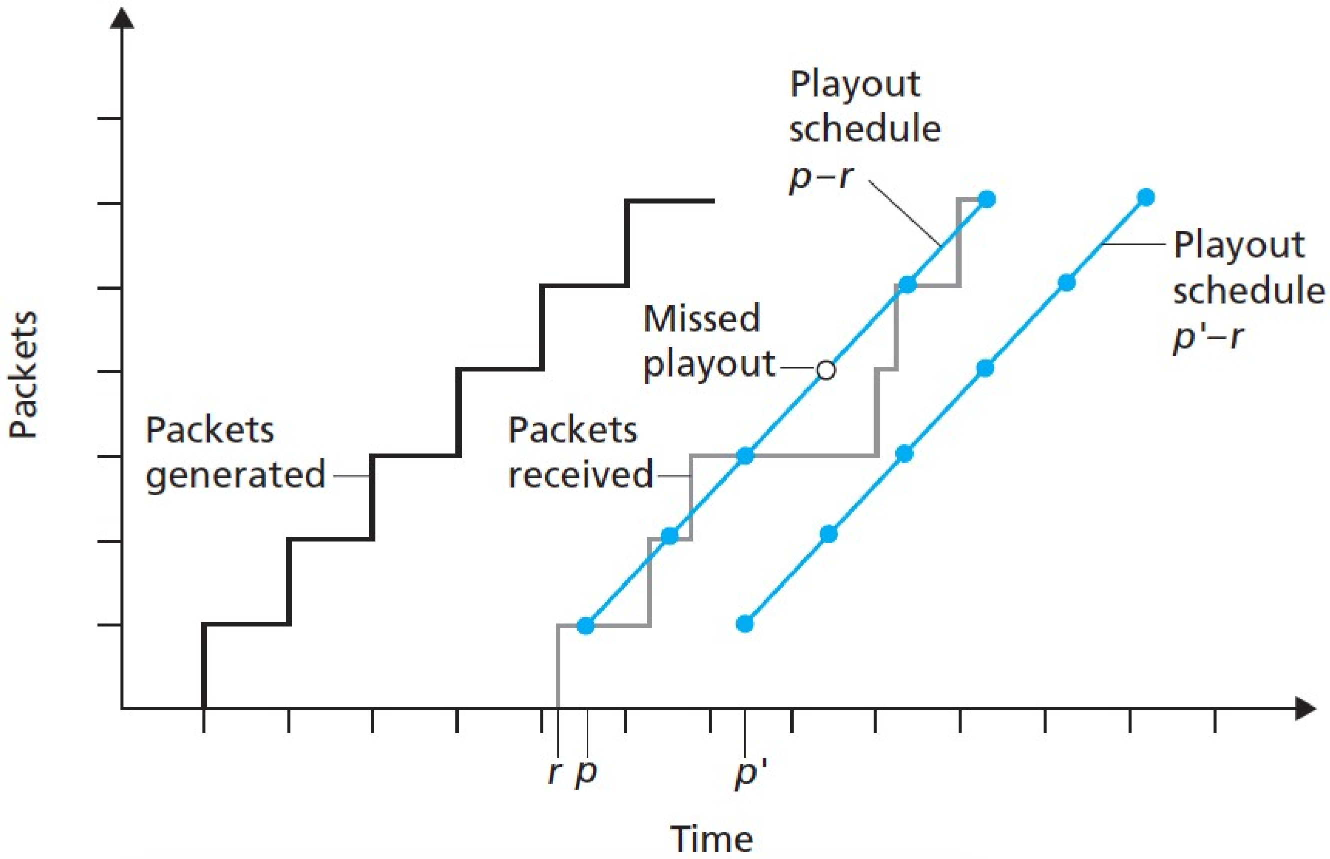
- **Packet Jitter**
- Varying queuing delay (per packet) at routers (on path) towards destination
- Some packets arrive in burst others in greater delay
- An earlier transmitted packet may arrive later while a packet transmitted later may arrive earlier (when transmitted on paths having different packet load)
- If the receiver plays out chunks as soon as they arrive, then the resulting audio quality can easily become unintelligible.



3. Conversational Voice and Video over Internet (VoIP: Removing Jitter at Receiver for audio)

- Fortunately, jitter can often be removed by using sequence numbers, timestamps , and a playout delay
- Timestamps tell when each chunk was generated and at what time gap to play it
- Delaying playout of chunks at the receiver, it can be either fixed or adaptive. It enable continuous audio reception.
  - **Fixed playout:** Each chunk is generated at t time, played at t+q time. Any delayed chunk is discarded. (has delay-loss trade-off). Long playout delay not good for conversation.
  - **Adaptive playout:** Estimate the network delay and the variance of the network delay, and adjusts the playout delay accordingly at the beginning of each talk spurt. So for each talk spurt the delay of playout may be different

3. Conversational Voice and Video over Internet (VoIP: Delayed Playout)



3. Conversational Voice and Video over Internet (VoIP: Removing Jitter at Receiver for audio)

- **Adaptive playout algorithm:**
- t_i: time when packet i is *generated* by sender
- r_i: when packet i is *received* by receiver
- p_i: when packet i is *played* by receiver
- d_i: **Average network delay** when packet i is received: $d_i=(1-u)d_{i-1} + u(ri-ti)$, where u=0.001
- vi: **average deviation of delay** from the estimated avg. delay: $vi=(1-u)v_i+u|ri-ti-di|$
- pi: **play out time for first packet** in talk spurt: $p_i=t_i+d_i+Kv_i$, where K=4
- $q_i=p_i-t_i$ (play out delay for packet i)
- Playout delay for every packet j in talk spurt with first packet **i** is:
- $p_j=t_j+q_i$

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

- A packet is lost either if it never arrives at the receiver or if it arrives after its scheduled playout time.
- Retransmission of a packet that has missed playout deadline serves no purpose.
- Loss recovery schemes used by VoIP
  - Forward Error Correction
  - Interleaving
  - Error Concealment

3. Conversational Voice and Video over Internet (VoIP: Removing Jitter at Receiver for audio)

- **Adaptive playout algorithm:**
- t_i: time when packet i is *generated* by sender
- r_i: when packet i is *received* by receiver
- p_i: when packet i is *played* by receiver
- d_i: **Average network delay** when packet i is received: $d_i=(1-u)d_{i-1} + u(ri-ti)$, where u=0.001
- vi: **average deviation of delay** from the estimated avg. delay: $vi=(1-u)v_i+u|ri-ti-di|$
- pi: **play out time for first packet** in talk spurt: $p_i=t_i+d_i+Kv_i$, where K=4
- $q_i=p_i-t_i$ (play out delay for packet i)
- Playout delay for every packet j in talk spurt with first packet **i** is:
- $p_j=t_j+q_i$

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

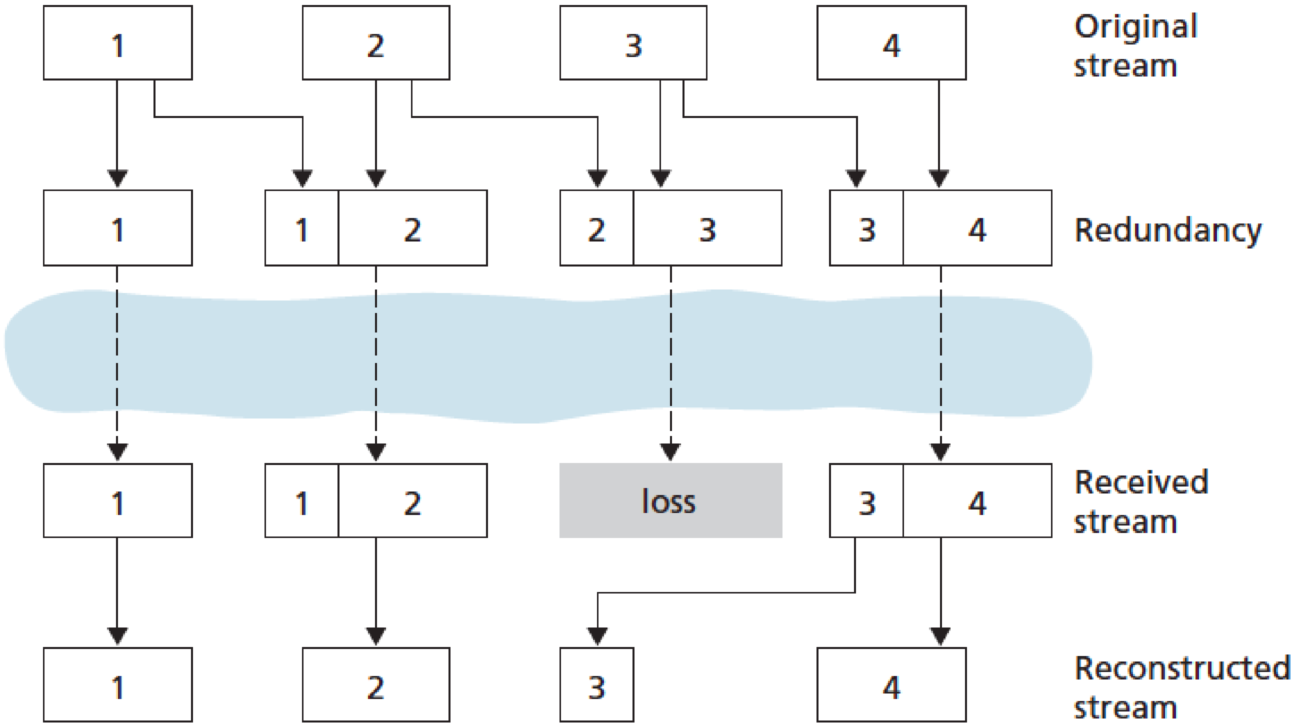
- **Forward Error Correction**
  - Add redundant information to the original packet stream
  - Redundant information helps in recovering lost information
  - **First Mechanism:**
    - Send a redundant chunk after every n chunks
    - The redundant chunk is generated by XORing the original n chunks
    - If one packet in the group of n+1 is lost, it can be fully recovered
    - If more than one packet is lost, loss can not be recovered
    - If n is small, larger number of lost packets can be recovered. The downside is transmission rate raise.
      - Example: if n=3, transmission rate raises by 1/3 or 33%
    - This scheme raises playout delay. Receiver has to wait more for entire group of packets to arrive.

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

- **Forward Error Correction**
  - **Second Mechanism:**
    - Send a a lower resolution stream as a redundant information.
    - Each n^{th} nominal bit-rate chunk is grouped with a low-bit rate $(n-1)^{th}$ chunk containing same information as earlier one but with lower quality.
    - When loss is non-consecutive, and a nominal bit-rate chunk is lost, the low-quality chunk is then played
    - Compared to the earlier mechanism, receiver has to wait for two chunk before playback, therefore playout delay is small.
    - If the low-quality has comparatively much smaller bit-rate then raise in transmission rate is marginal.
    - To recover from consecutive loss:
      - Each $(n)^{th}$ nominal chunk be grouped with
      - (n-1) and (n-2) OR (n-1) and (n-3) low bit rate chunk

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

- **Forward Error Correction**
- Piggybacking lower quality redundant information



3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

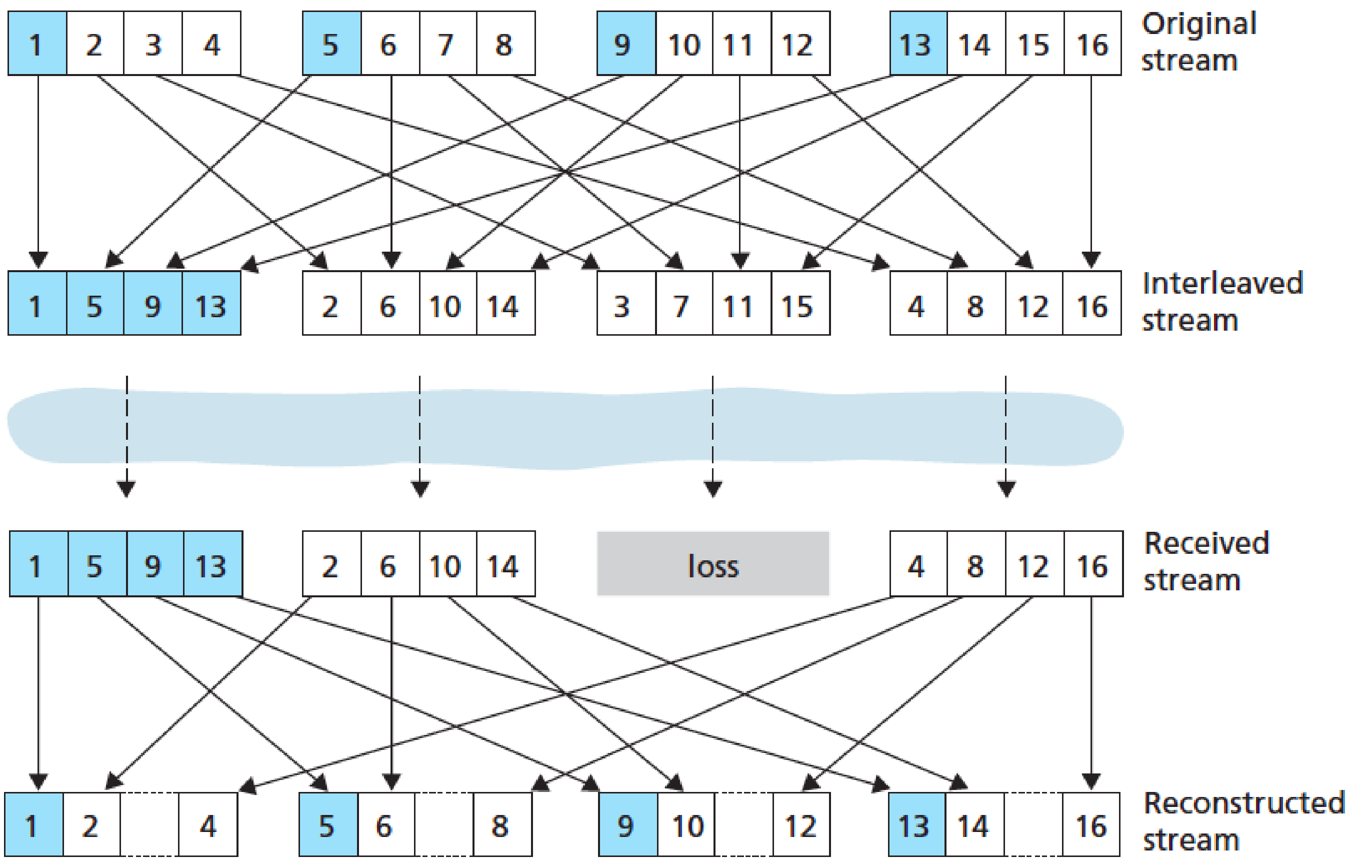
- **Interleaving**
  - The original consecutive units are separated by a distance in transmission stream
  - Mitigates effect of packet loss
  - Example:
    - Each unit is of 5 msec
    - Each chunk is of 20 msec
    - Then each chunk has 4 units. First chunk will have unit 1,2,3 and 4 and so on.
    - But in transmission first chunk will carry units 1,5,9 and 13 and
    - Second chunk will carry units 2,6,10 and 14 and so on.

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

- **Interleaving**
- The loss of a single packet from an interleaved stream results in multiple small gaps in the reconstructed stream, as opposed to the single large gap that would occur in a non-interleaved stream.
- Improves the perceived quality of an audio stream
- It also has low overhead, does not raise bandwidth requirement.
- Disadvantage of interleaving is that it increases latency
  - Limits its use for VoIP
  - BUT can be used for streaming stored video/audio

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

- **Interleaving**



School of Informatics
Institute for Computing Systems Architecture

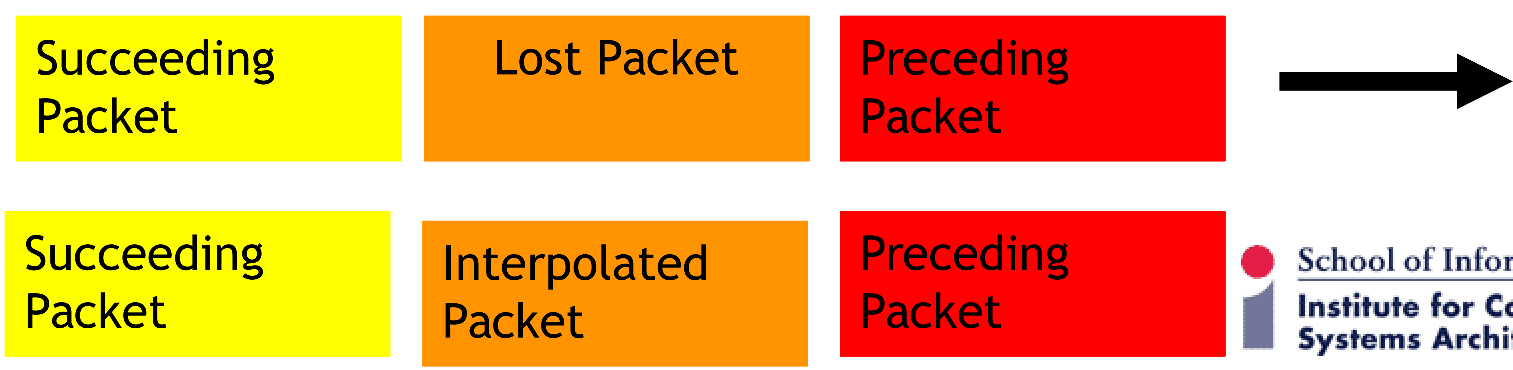
3. Conversational Voice and Video over Internet
(VoIP: Recovering from packet Loss)

- **Error Concealment (Packet Repetition):**
- Attempt to produce a replacement for a lost packet that is similar to the original
- Possible since audio signals, and in particular speech, exhibit large amounts of short-term self-similarity.
- Works for relatively small loss rate (< 15%) and for small packet (4-40msec)
- Packet repetition with packets arriving immediately before the loss.
- Low computational complexity and performs reasonably well.

Succeeding Packet	Lost Packet	Preceding Packet	→
Succeeding Packet	Packet Repetition	Preceding Packet	

3. Conversational Voice and Video over Internet (VoIP: Recovering from packet Loss)

- **Error Concealment (Interpolation):**
- Uses audio before and after to interpolate (estimate) the information in the lost packet.
- Performs better than packet repetition BUT somewhat computationally expensive



3. Conversational Voice and Video over Internet (Case Study: Skype)

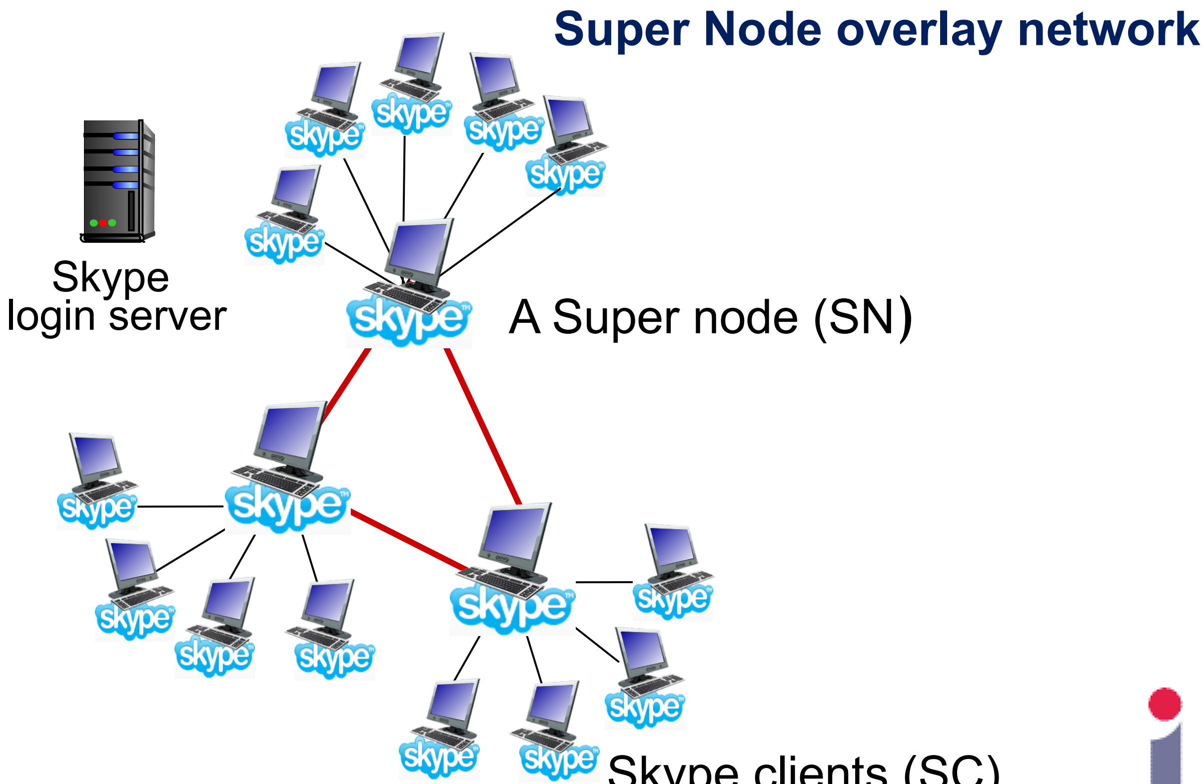
- Supports a variety of conversational voice/video over IP services: host-to-host VoIP, host-to-phone, phone-to-host, multi-party host-to-host video conferencing
- A host is again any Internet connected IP device (including PCs, tablets, and smartphones.)
- A property of Microsoft
- A proprietary system:
  - Encrypted control and media messages, so one is unable to see how it operates
  - what we know about it is mainly via reverse engineering by researchers using measurements

3. Conversational Voice and Video over Internet (Case Study: Skype)

- Uses wide variety of audio and video codecs
  - Offering different audio/video rates and qualities
  - Video quality ranges from 30kpbs to 1Mbps
  - Audio codecs use a higher sampling rate 16000 samples/second.
    - Plain old telephone system (POTS) provides 8000 samples/sec
    - Skype audio quality is usually better than POTS
- Control packets over TCP
- By default, audio and video packets sent over UDP
  - Uses FEC for loss recovery
- Adapts audio and video streams to current network conditions by varying audio/video bit-rate and FEC overhead

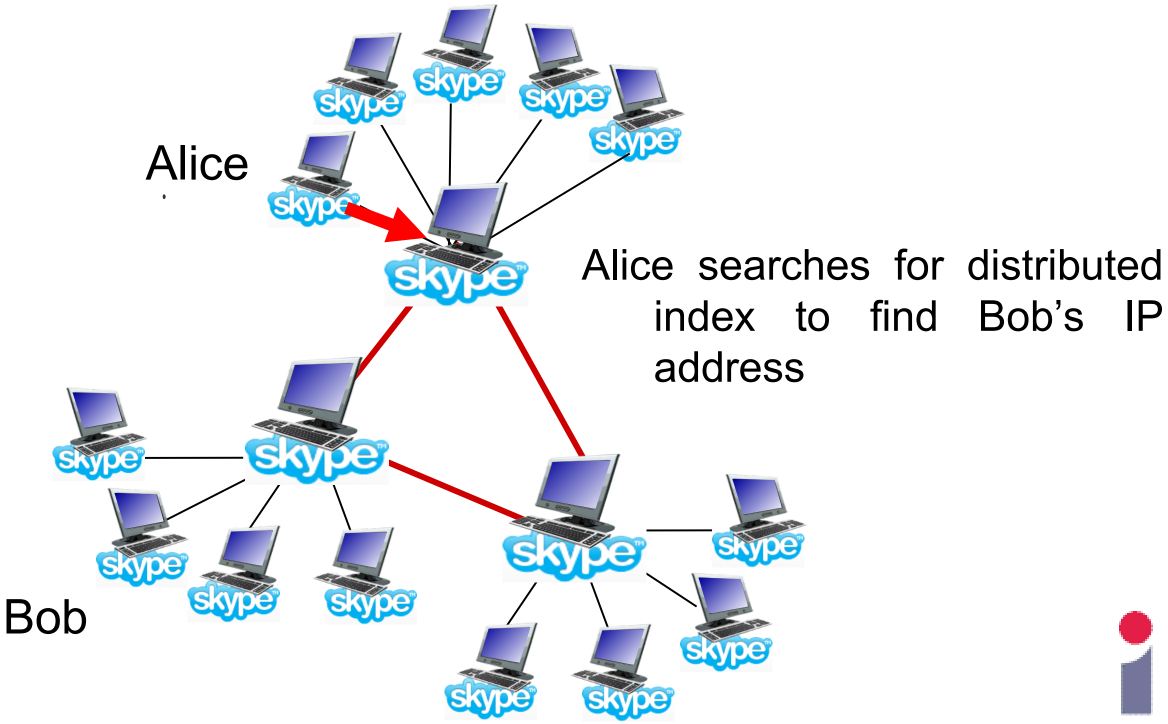
3. Conversational Voice and Video over Internet (Case Study: Skype has an overlay network)

- A hierarchical overlay network with
  - Super nodes (SNs)/ Super Peers
  - and Skype clients / client peers



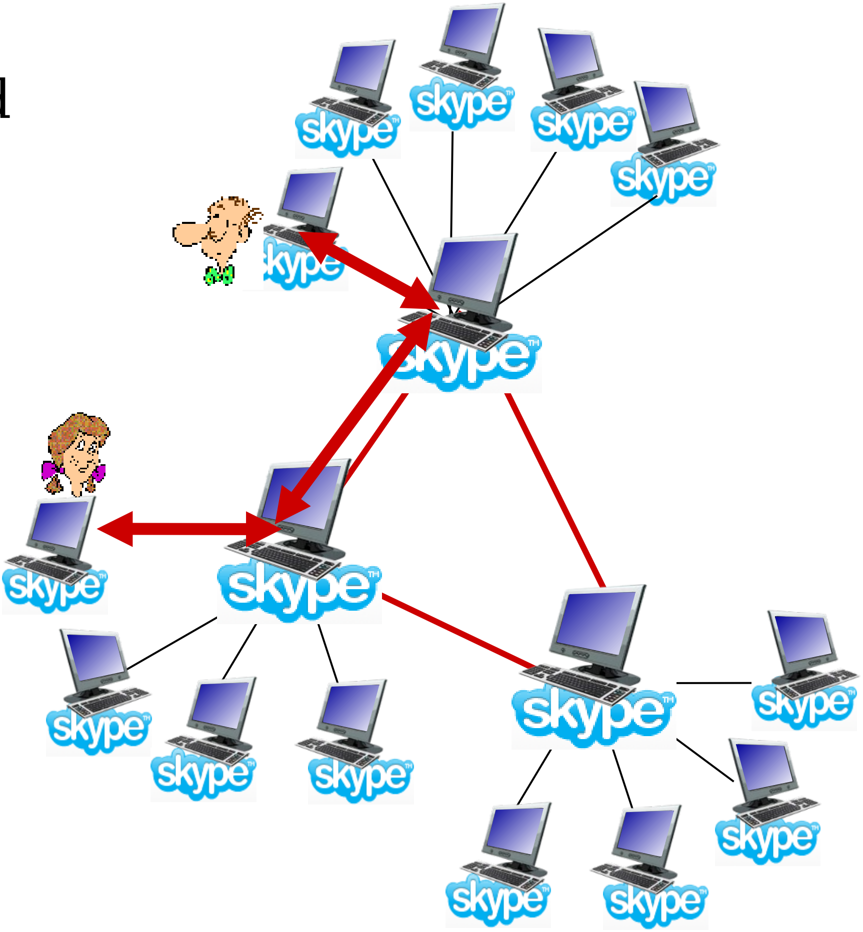
3. Conversational Voice and Video over Internet (Case Study: Skype A P2P system)

- **User location:**
- Skype maintains an distributed index that maps Skype usernames to current IP addresses (and port numbers).
- Distributed Hash Table (DHT) organization is very possible at Super peers.



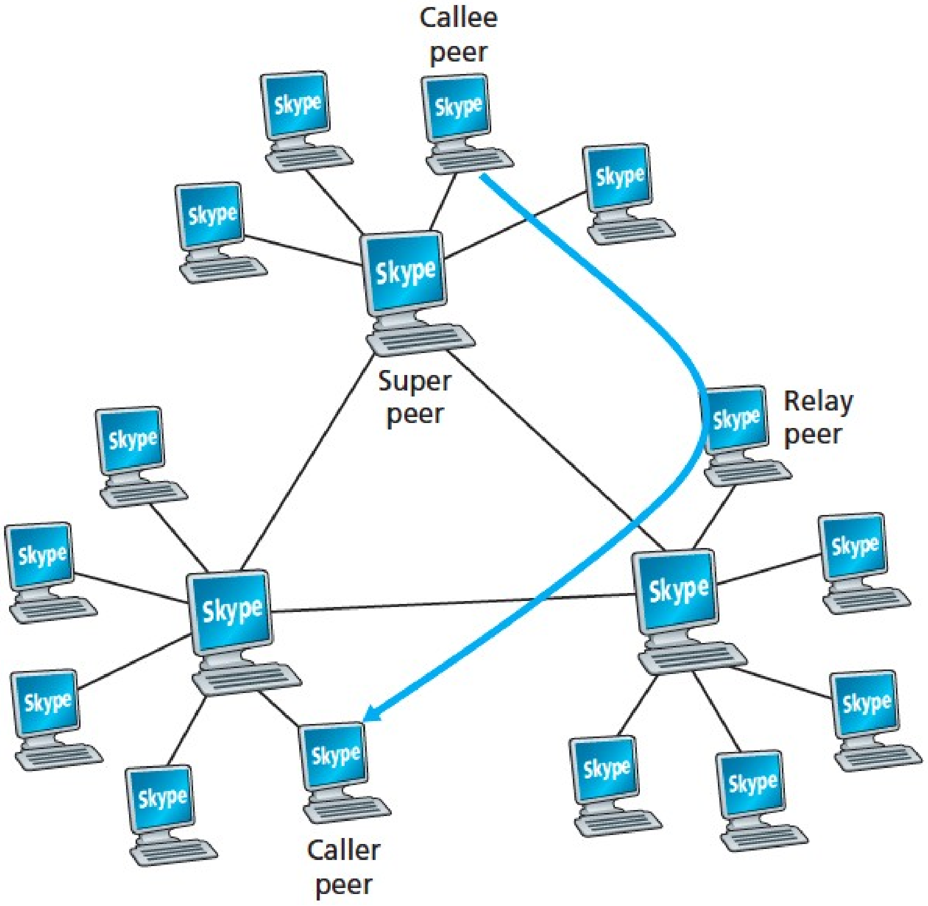
3. Conversational Voice and Video over Internet
(Case Study: Skype A P2P system)

- **NAT Traversal:**
- Problem: if both parties behind NATs
  - NAT prevents outside peer from initiating connection to a peer behind it
  - Peer behind NAT *can* however initiate connection to outside
- Super peers can enable connection between client peers (that are behind NAT)



3. Conversational Voice and Video over Internet
(Case Study: Skype A P2P system)

- **Host-to-host Internet telephony** is inherently P2P
- SN as a relay
  - Client maintain open connections to their respective SNs
  - Caller client notifies its SN to connect to callee
  - SNs for caller and callee connect with each other
  - Another SN is identified as a relay between the two parties and both connect to that SN and engage in the call



3. Conversational Voice and Video over Internet
(Case Study: Multi-part Skype conferences)

- Multi-Party = number of participants, ***N* > 2**
- With audio streams, the conference initiator collects streams from all parties, combines them and distributes back the combined stream
  - reduce communication overhead from sending *N(N-1)* to ***2(N-1)*** streams
- With video streams, each participant's video stream sent to a server cluster which in turn sends to the rest of the participants
  - *N(N-1)* streams sent in total but does not stress the upstream access links which typically have much lower bandwidth than downstream links