Real Time Communications

- In most data communications, important that data arrives reliably
  - Would like it to be fast, but prefer reliable
    - E.g. web, email, p2p, etc.
  - Often characterised as *elastic* applications

- In real time communications it is important that the message arrives in a timely manner
  - Timeliness may be *more* important than reliability
  - Messages may have priority
  - Examples:
    - A "drive by wire" system in a car
    - Packet voice and telephony applications

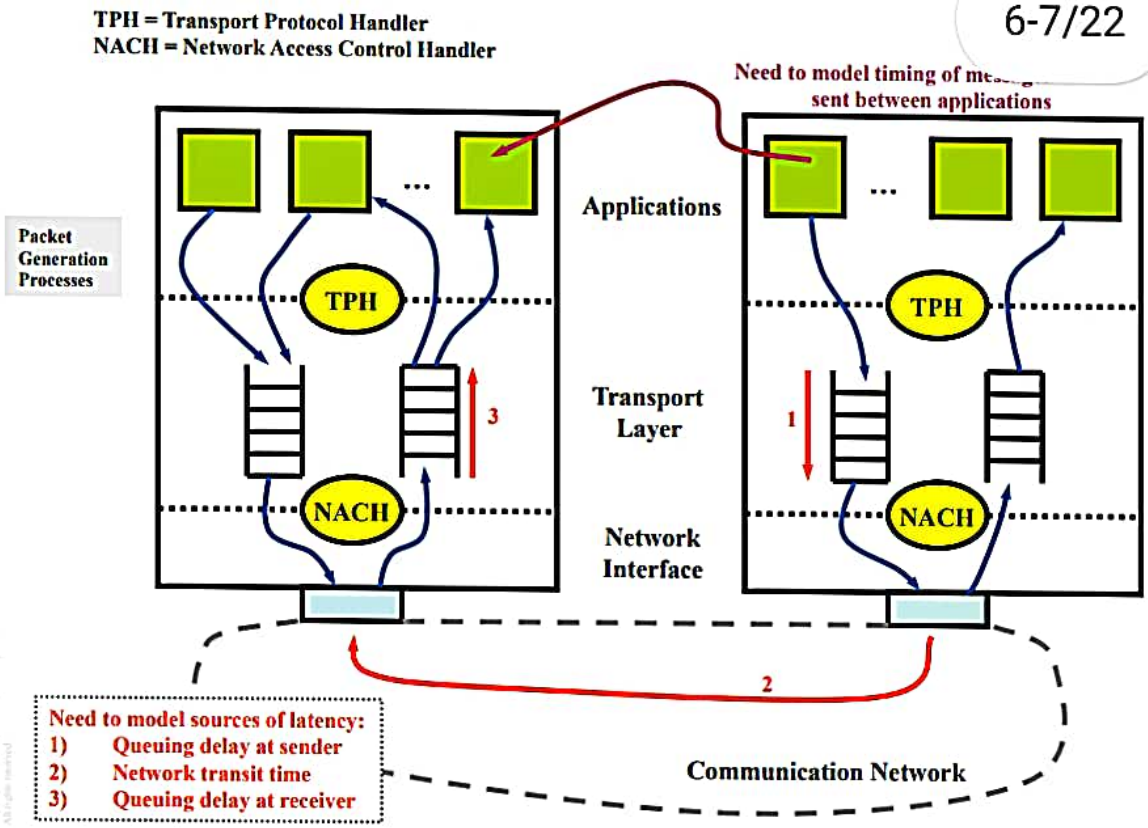
Modelling Real Time Traffic

- Assume a packet-based network
  - Real-time traffic on circuit switched network trivial after connection setup
- Traffic falls into two categories:
  - Synchronous periodic flows
    - Produced and consumed in a continuous basis, according to some schedule
    - Generally require some performance guarantee
    - Can be generated by periodic tasks
      - Fixed rate ("isochronous") flows (e.g. sensor data, speech)
      - Characterise by inter-packet spacing, message length, reception deadline
    - Can be generated by sporadic tasks
      - Variable rate flows (e.g. MPEG-2 video, control traffic)
      - Characterise by average throughput + maximum burst size
  - Aperiodic (asynchronous) messages
    - No deadline, best effort delivery, but want to keep delays small
    - Characterise by average delivery time

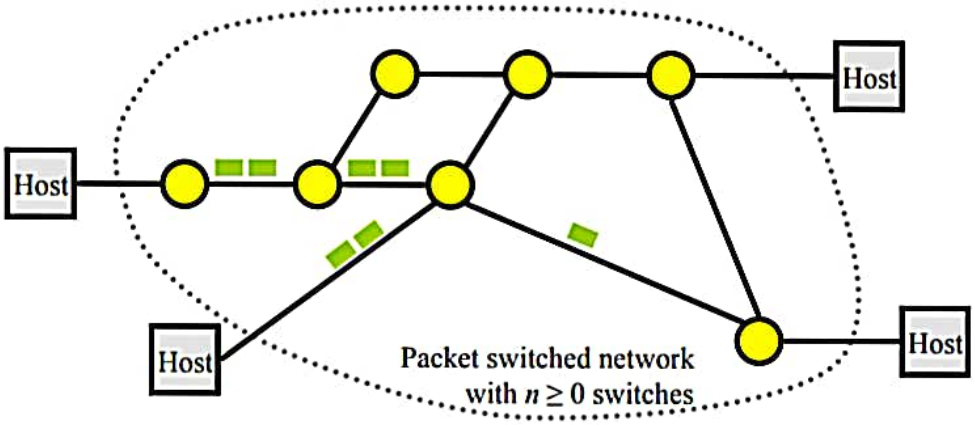
Modelling Sources of Timing Variation

- Ideally the network delivers messages to receiver with no delay, preserving timing
- In reality there is:
  - Queuing delay at sender
    - Network not always ready to accept a packet when it becomes available; data may be queued if produced faster than the network can deliver it
  - Queuing delay at receiver
    - Application not always ready to accept packets arriving from network
    - Network may deliver data in bursts
  - Queuing delay in the network
    - Due to cross-traffic or bottleneck links
  - Network transit time
    - Fixed propagation delay

Model of Hosts



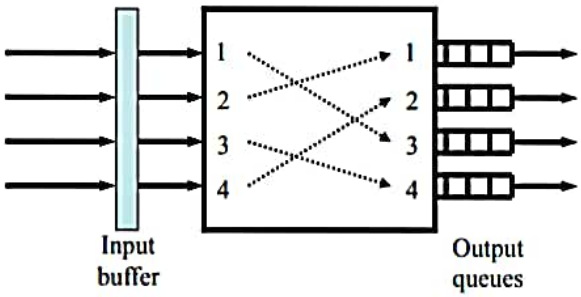
Model of Packet Switched Networks



Links have constant *propagation delay*

Switches queue packets for transmission if output link busy (additional variable delay)

Choice of *job scheduling algorithm* on the output link is critical for real time traffic

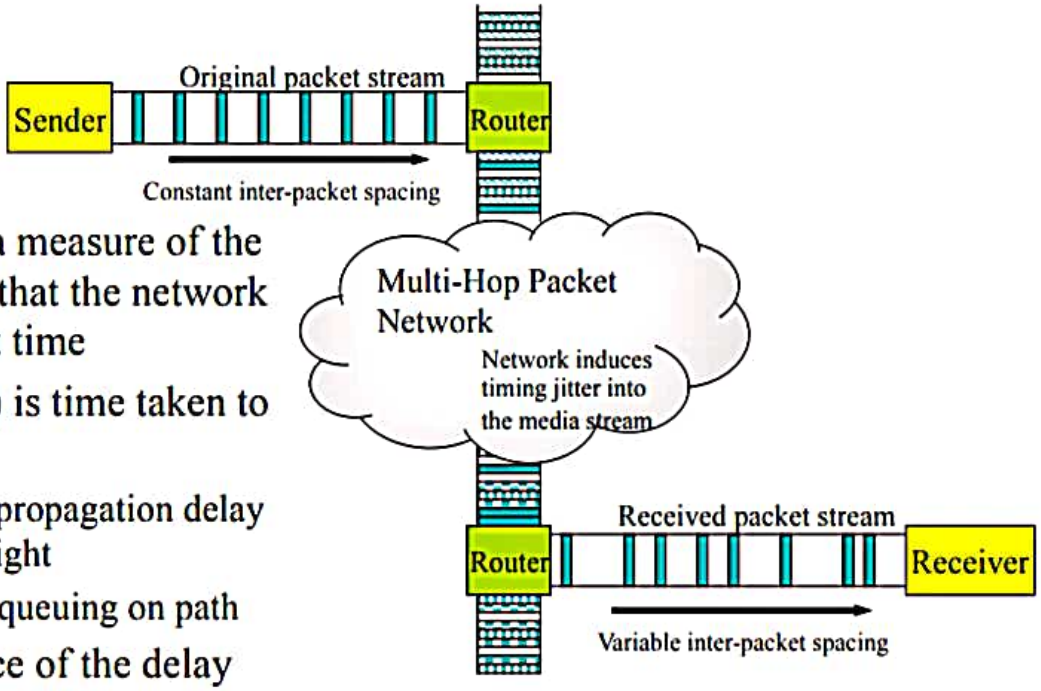


Performance Metrics and Constraints

- From these models, derive performance metrics:
  - Throughput and delay
  - Jitter and buffer requirements
  - Miss rates, when jitter causes a deadline to be missed
  - Packet loss and invalid rates

- Characterise traffic and network according to metrics to schedule communications
  - Need to meet application timing constraints

Throughput, Delay and Jitter



- The *throughput* is a measure of the number of packets that the network can deliver per unit time
- The *delay* (latency) is time taken to deliver a packet
  - Fixed minimum propagation delay due to speed of light
  - Variation due to queuing on path
- The *jitter* is variance of the delay

- Throughput, delay and jitter vary according to router scheduling algorithms
  - Possible to derive bounds for delay/jitter in some cases
  - Lecture 16

Example: Ethernet

- Recall that Ethernet uses CSMA/CD with exponential back off
  - Try to transmit, listening for collision
  - If a collision occurs, stop sending, wait before retry
  - Random binary exponential back-off
    - After i collisions back-off by up to $2i$ slots, randomly chosen
- Potentially unbounded delay on busy network
  - Cannot schedule transmissions to avoid collision
- No prioritisation of messages

- Implications:
  - Cannot easily reason about timing properties
  - Difficult to schedule messages to ensure timely delivery

Summary

- What is real time communication
- Factors that affect real time communication
  - Throughput, delay and jitter
  - Clock skew
  - Congestion and loss
- Examples of networks and their timing properties
  - Some networks provide timing guarantees, others do not