Real-Time Communication

Sathish Gopalakrishnan
Electrical and Computer Engineering
The University of British Columbia

Outline

- Real-time communications
  - Traffic and network models
  - Properties of networks
    - Throughput, delay and jitter
  - Congestion and loss
- Examples
  - Controller area networks
  - Ethernet

Real-Time Communications

- Digital data communications
  - One would like fast delivery, but not at the expense of reliability
    - E.g. web browsing, e-mail, file transfer, twitter, etc.
  - These applications are often referred to as elastic applications
    - i.e. time can be “dilated”

- Real-time data communications
  - Timely delivery may deemed to be more desirable than reliable delivery
  - Different levels of priority may be associated with applications
  - Examples:
    - Anti-lock braking in a car
    - “Fly-by-wire” systems in a modern aircraft
    - Skype internet telephony and IPTV (TV using the Internet Protocol)
      - drop delayed packets

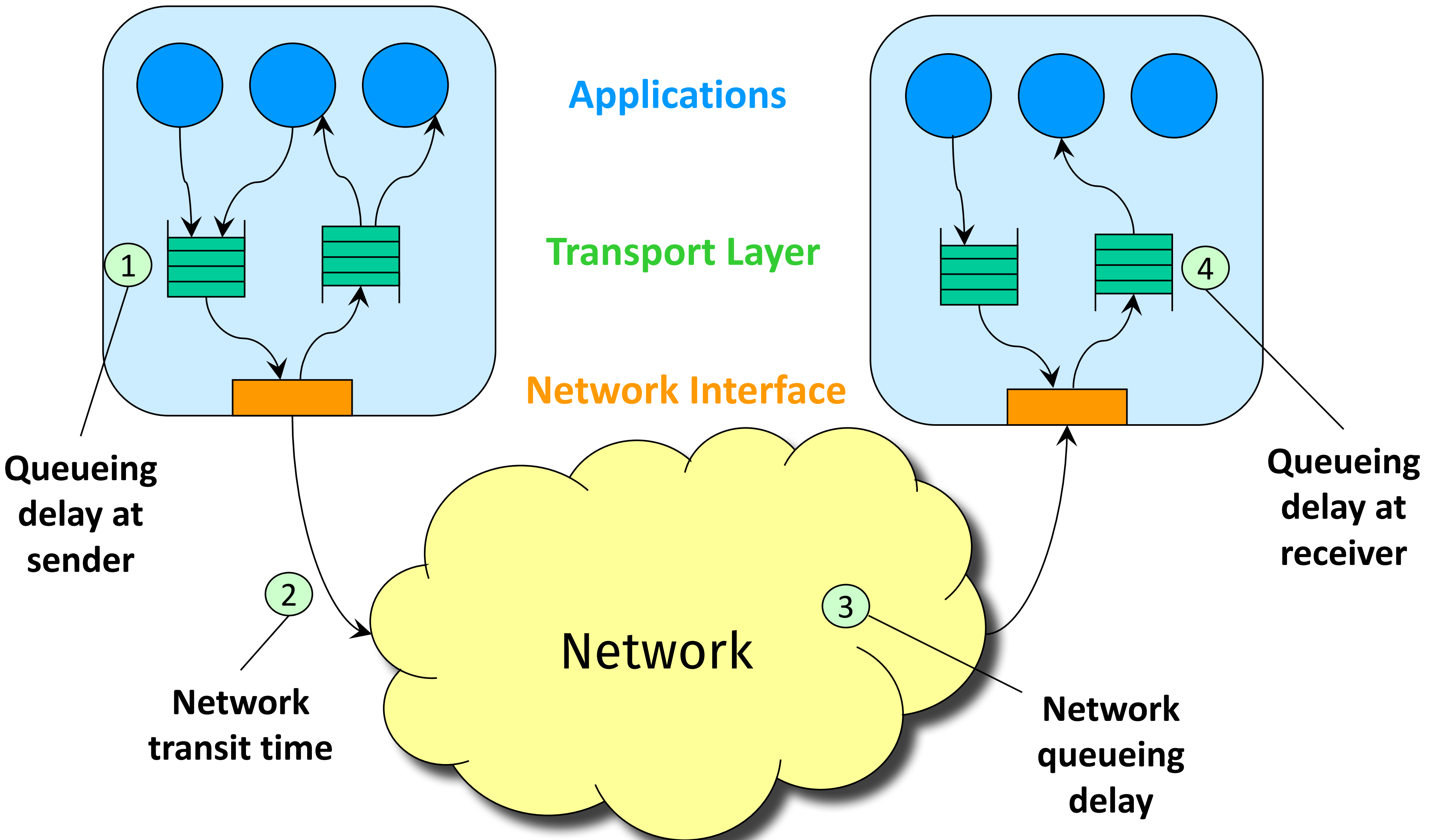
Real-Time Traffic Categories

- Packet-switched traffic falls into two categories:
  - **Synchronous periodic messages**
    - Produced and consumed in a continual 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)
      - Characterize by inter-packet spacing, message length, reception deadline
    - Can be generated by sporadic tasks
      - Variable rate flows (e.g. MPEG-2 video, control traffic)
      - Characterize by average throughput + maximum burst size
  - **Aperiodic (asynchronous) messages**
    - No deadline, best-effort delivery, but want to keep delays small
    - Characterize by average delivery time

Sources of Message Delays

- Message delays on networks comprise the following components:
  1. 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
  2. Queuing delay in the network
     - Due to cross-traffic or bottleneck links
  3. Network transit time
     - Fixed propagation delay
  4. Queuing delay at receiver
     - Application not always ready to accept packets arriving from network
     - Network may deliver data in bursts

Network Message Delay



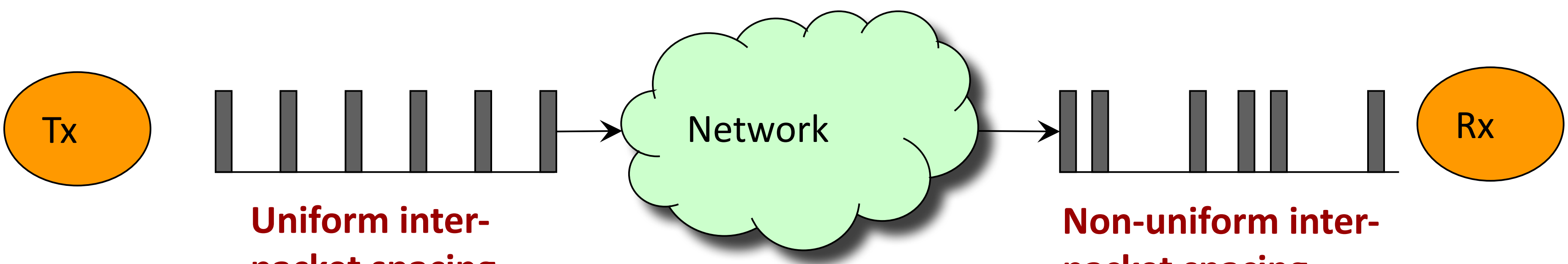
Performance Metrics

- **Throughput**: a measure of the number of packets that the network can deliver per unit time
- **Delay (latency):** time taken to deliver a packet
  - Fixed minimum propagation delay due to speed of light
  - Variation due to queuing on path
- **Jitter:** Variance of the delay
- **Buffer requirements:** amount of storage required so as not to drop packets
- **Packet Miss rate:** ratio of packets that miss their timing constraints
- **Packet Loss Rate:** ratio of packets that are not delivered
- **Packet Error Rate:** ratio of packets that have an error in them

Throughput, Delay and Jitter

Throughput, delay and jitter vary according to router scheduling techniques.

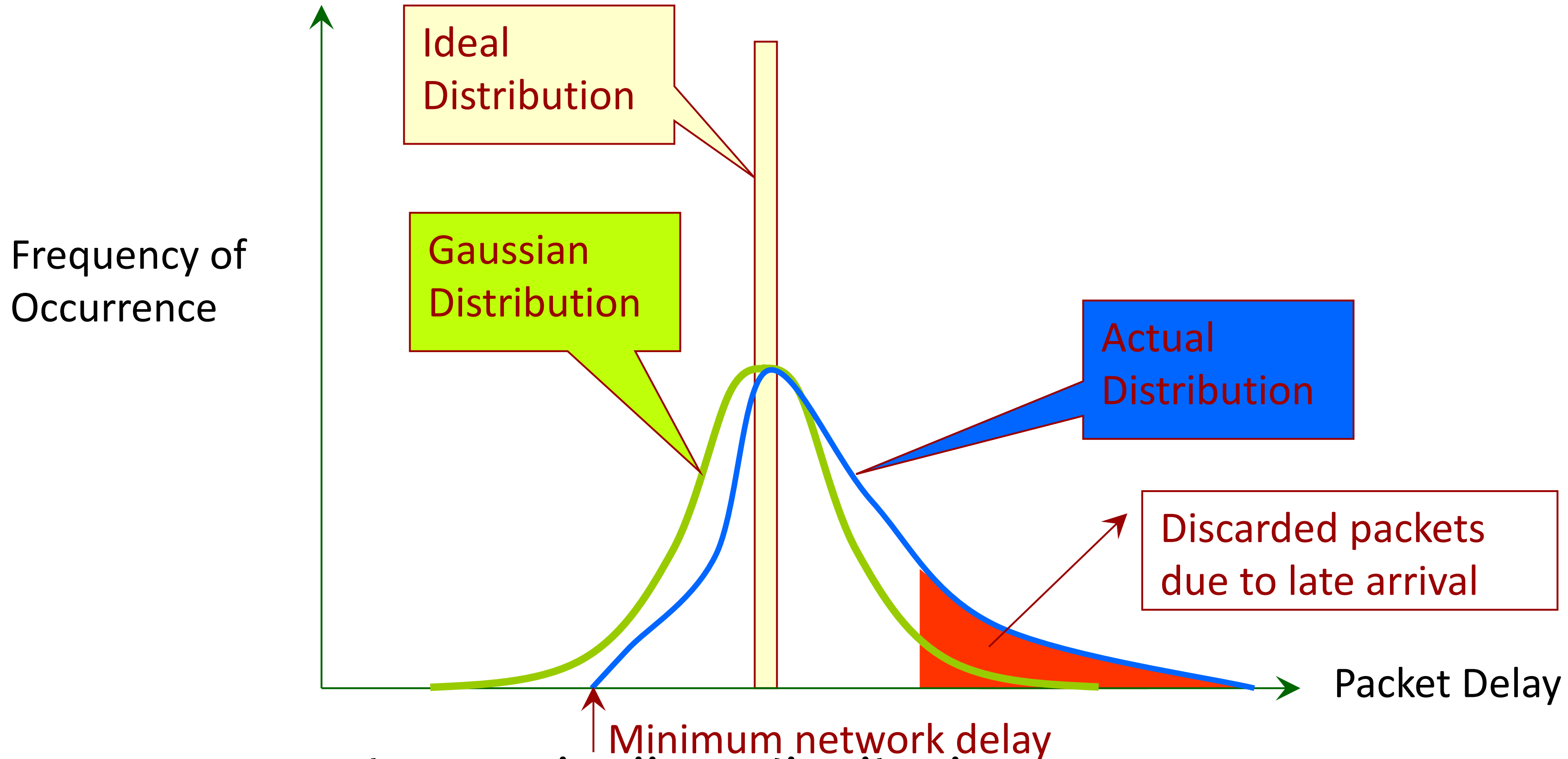
- Possible to derive bounds for delay/jitter for some policies (e.g. RMA, Round-Robin, Weighted Round-Robin techniques)



Uniform inter-packet spacing

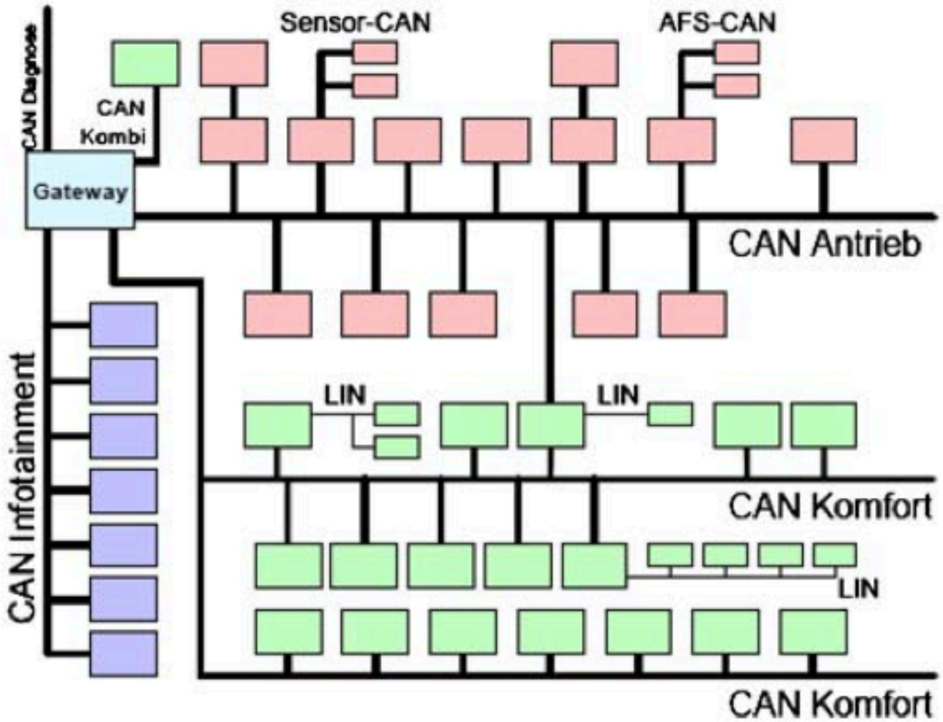
Non-uniform inter-packet spacing
(some packets delayed more than others)

Jitter and Miss-Rate



- Want to characterize jitter distribution
  - Hope for something approximating a Gaussian distribution ⇒ simple statistics to derive the miss rate
    - Fraction of packets lost due to jitter
  - Actual distribution is more complex

VW Passat Network Architecture



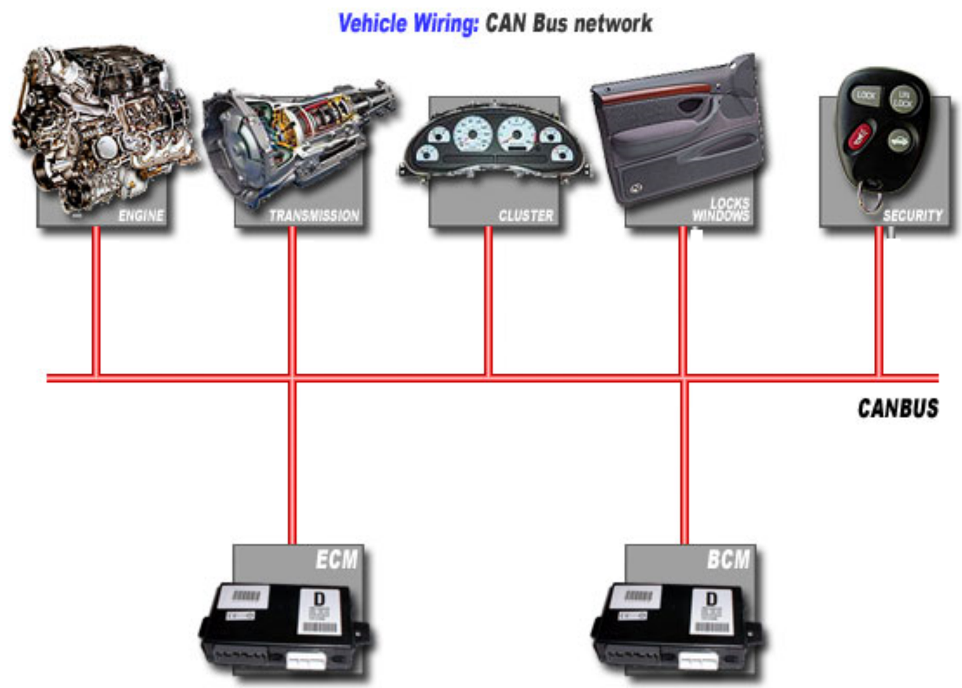
Often we find many different physical / MAC layers within a system…

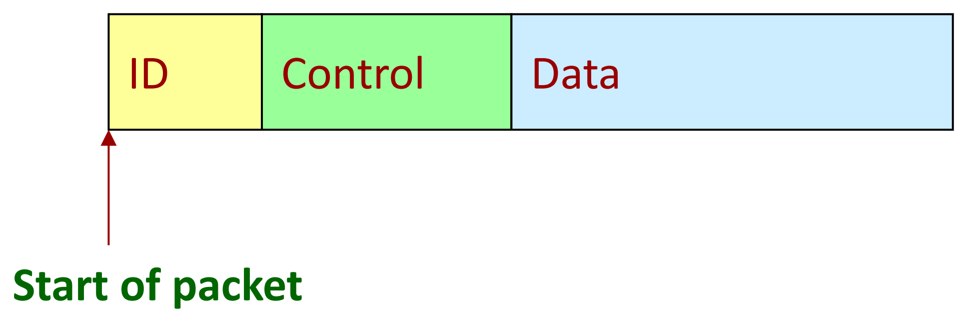
Congestion and Loss

- Both flow characteristics and cross-traffic can cause overloads and congestion, so might the cross traffic
  - Temporary congestion will cause queuing delays
  - Persistent congestion will result in queues that stay full, hence packets may be lost
- How to avoid this?
  - Control the amount of traffic at a bottleneck link
    - Applications need to signal their requirements
    - Network performs admission control
  - Or prioritize traffic to give preference to important flows
    - What scheduling algorithm to use?
      - Fixed-priority schemes are much easier to implement than dynamic priority schemes
      - Weighted round-robin techniques are also available
    - May allow real-time traffic, but discard best-effort data traffic when the network is overloaded

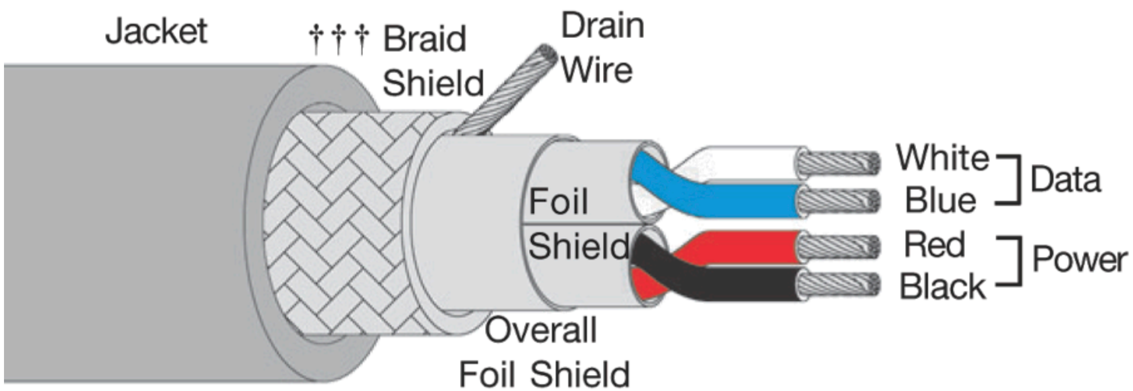
Controller Area Networks (CANbus)

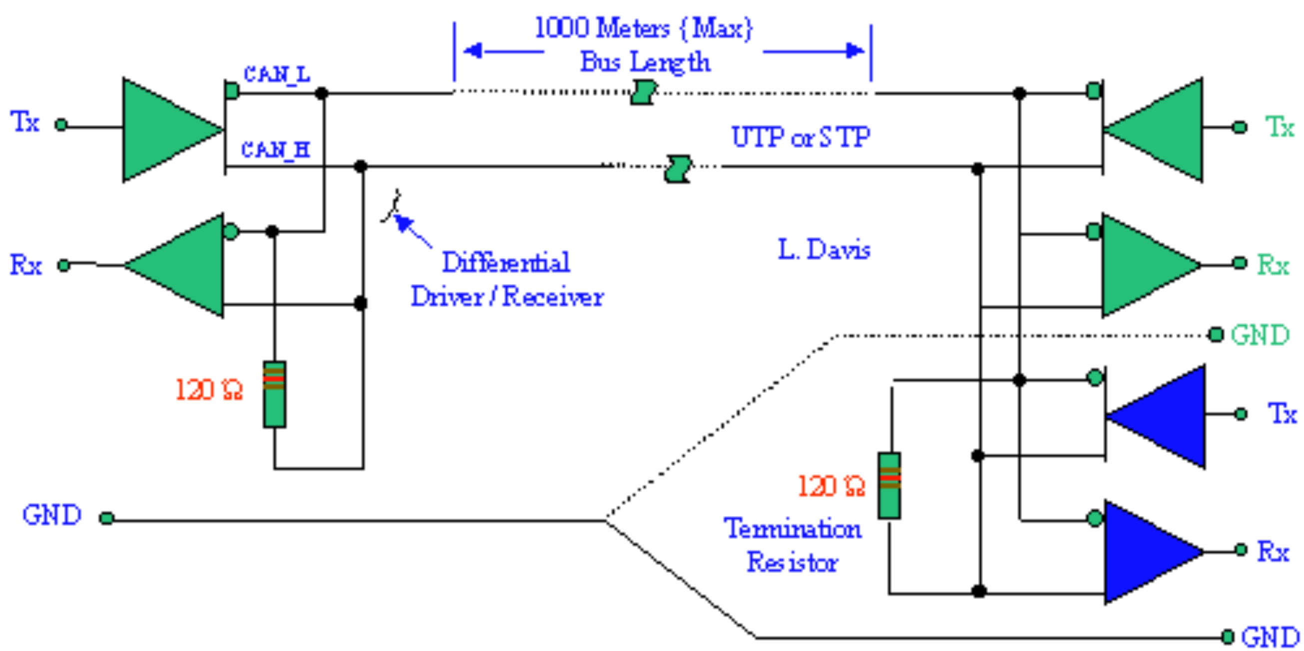
- Shared serial bus, send at 1Mbps, maximum bus length is 50m
- All stations hear transmissions within a fraction of a bit time
- Connections **wired AND** logic (zero will dominate) logic
- Packets start with an ID, then control and data
- Widely used in automotive
  - Finding its way to other areas



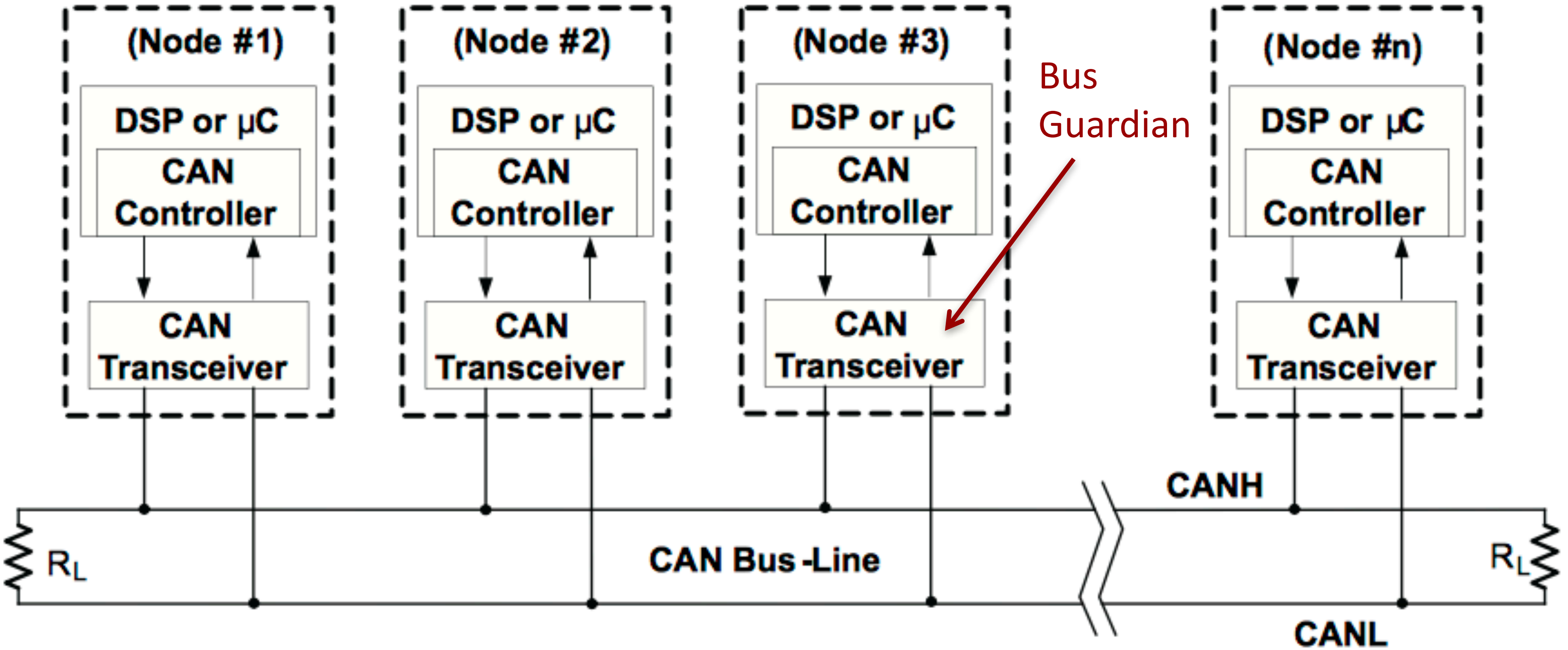


CAN bus: Wiring



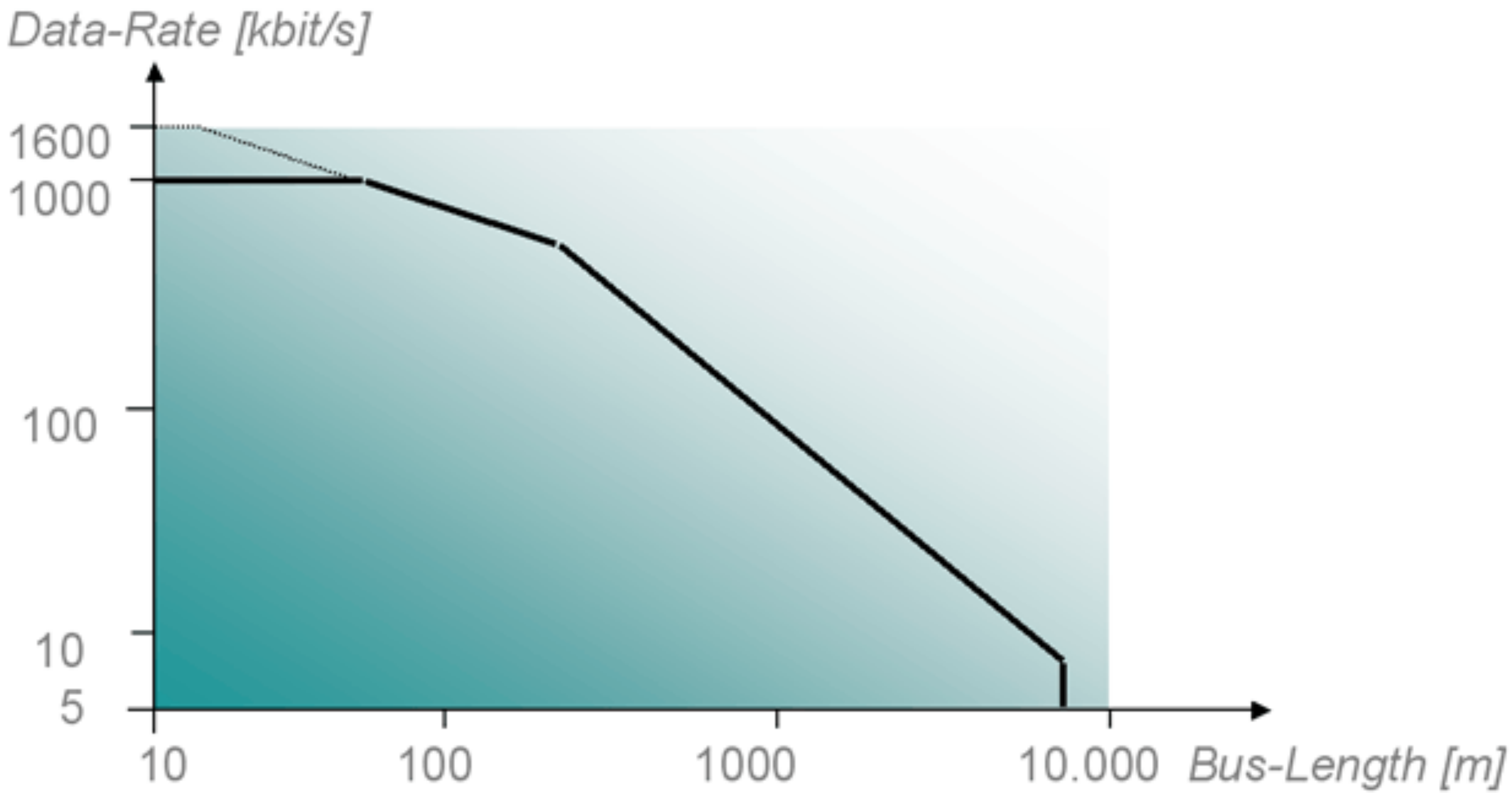


CAN Bus



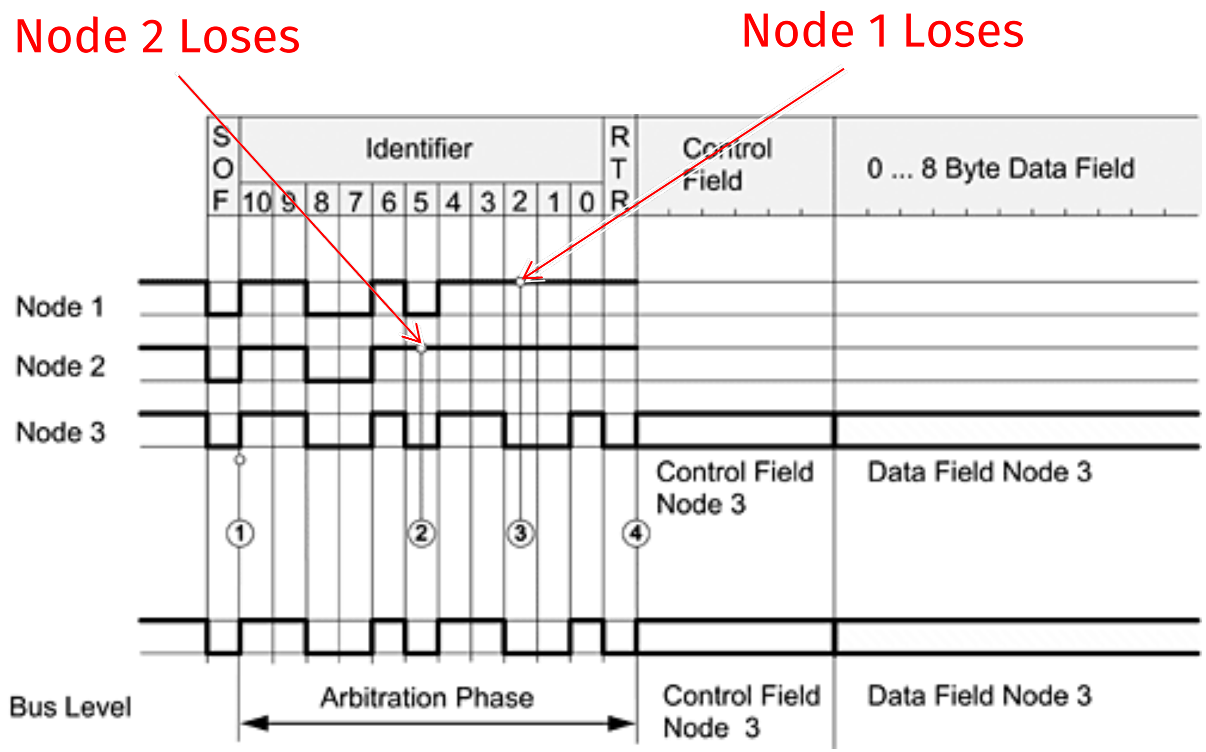
Courtesy of TI sloa101 application note…

CAN bus: Data Rate vs Distance

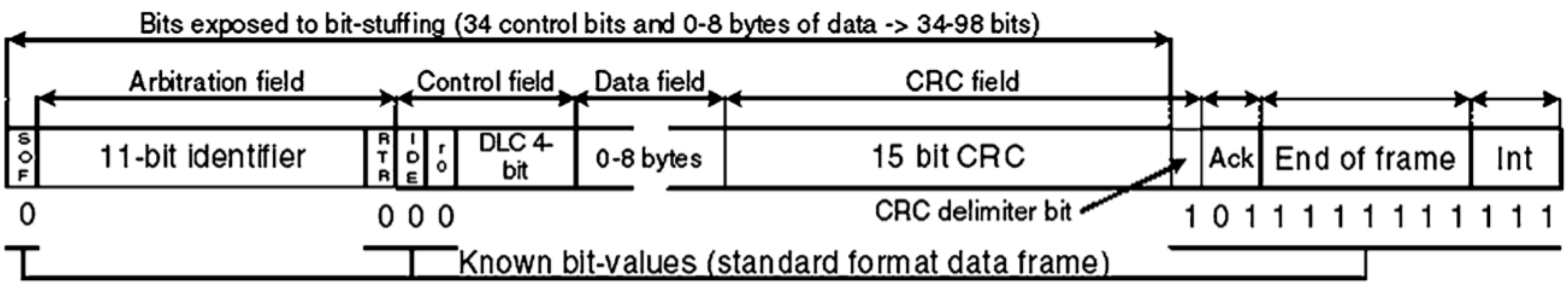


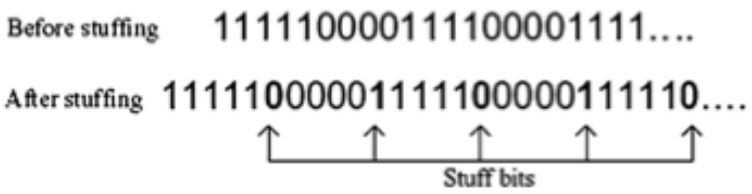
Distributed Priority-based Arbitration

- Wait until start of slot, then begin to send with the ID field, but:
- Stop if you hear a 0 on the bus when you are sending a 1
- Packet with the smallest ID is transmitted first
  - Priority-based network access



CAN bus: Packets





Real-Time Analysis of CAN

Remember our CPU task model $\tau_i = \{ C_i, T_i \}$

- In a communication system:
  - C_i is the message transmission time
  - T_i is the period of the message

- What is different as compared to a CPU?

$$\frac{C_1 + B_1}{T_1} + \ldots + \frac{C_n + B_n}{T_n} \leq U(n) = n(2^{1/n} - 1)$$

CAN bus Timing Analysis

- Treat messages like priority-based tasks on a processor
  - Perform RMA analysis
- Messages have non-preemptible regions
  - Adjust by adding blocking factors
- Davis, et al “Controller Area Network (CAN) schedulability analysis: Refuted, revisited and revised” RTSJ, 2007

Example: Ethernet

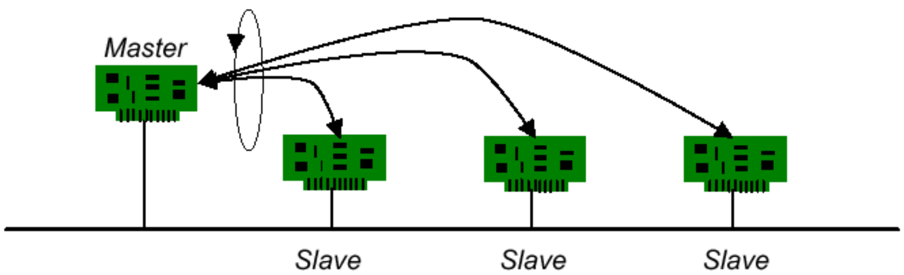
- Recall that Ethernet uses CSMA/CD with exponential back off
  - Before transmitting, check for active link
  - If not active, try to transmit, listening for collision
  - If a collision occurs, stop sending, wait before retry
  - Random binary exponential back-off
    - After *i* collisions, back-off by up to 2*i* slots, randomly chosen

- Potentially unbounded delay on busy network
  - Cannot schedule transmissions to avoid collision
- No prioritization of messages

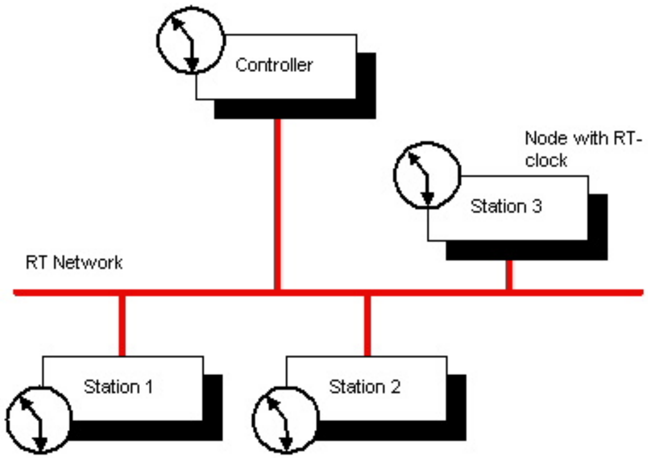
- Implications:
  - Throughput actually drops at high loads
  - Cannot easily reason about timing properties
  - Difficult to schedule messages to ensure timely delivery

Real-Time Ethernet?

- If you have control over ALL nodes in the system
- Coordinate using Time Diversity Multiple Access (TDMA)
- Synchronize each node (somehow) and assigned a time slot to transmit



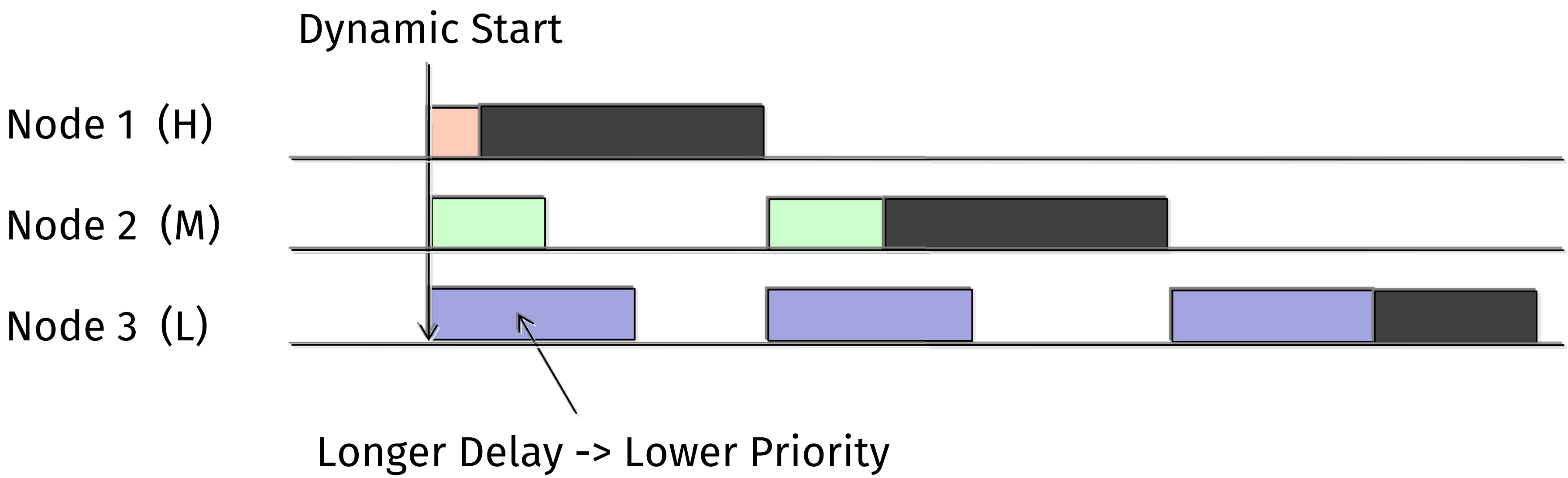
Polling TDMA



Synchronized TDMA

Priorities in non-Logical AND networks?

- Use transmission delay to act like a priority
- Low priority waits longer before it transmits



Conclusions

- What is real-time communications?
- Factors that affect real-time communication
  - Throughput, delay and jitter
  - Clock skew
  - Congestion and loss
- Examples of networks and their timing properties
  - Some networks (like CANbus) provide timing guarantees, others (like the ethernet) do not