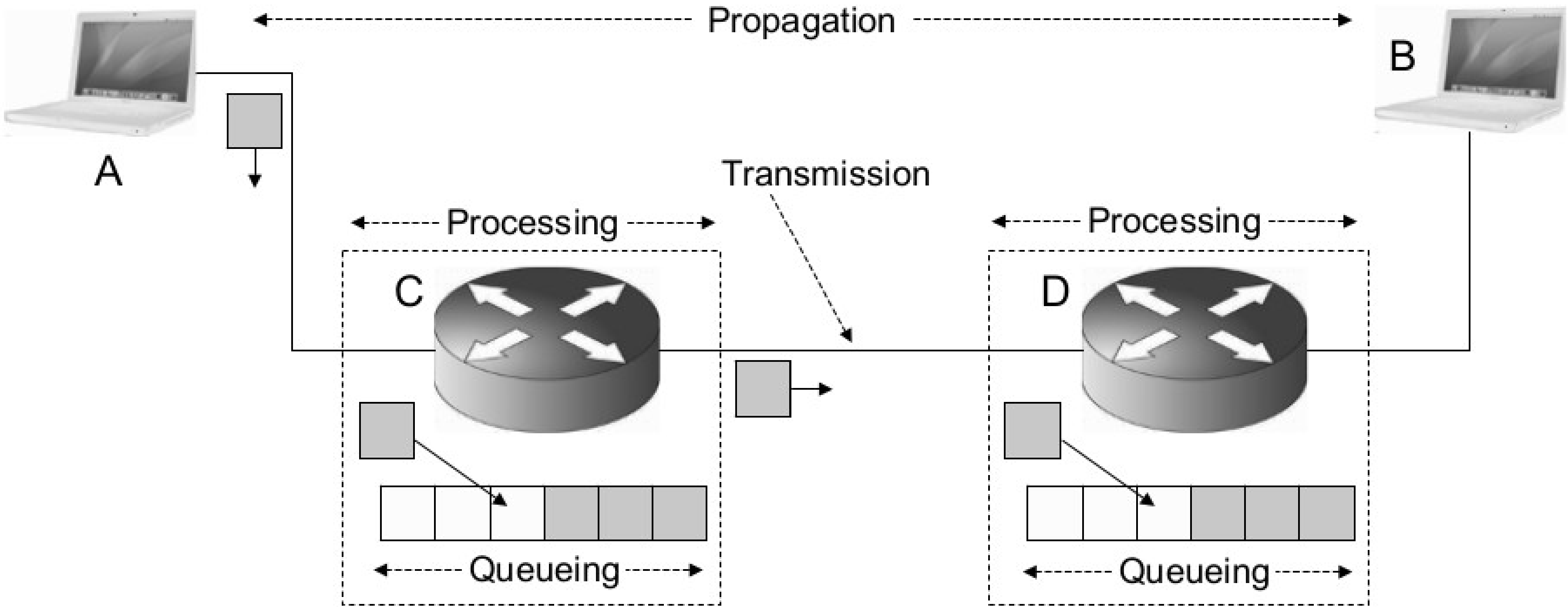
Lecture 9

Communication Fundamentals in Computer Networks

M. Adnan Quaium

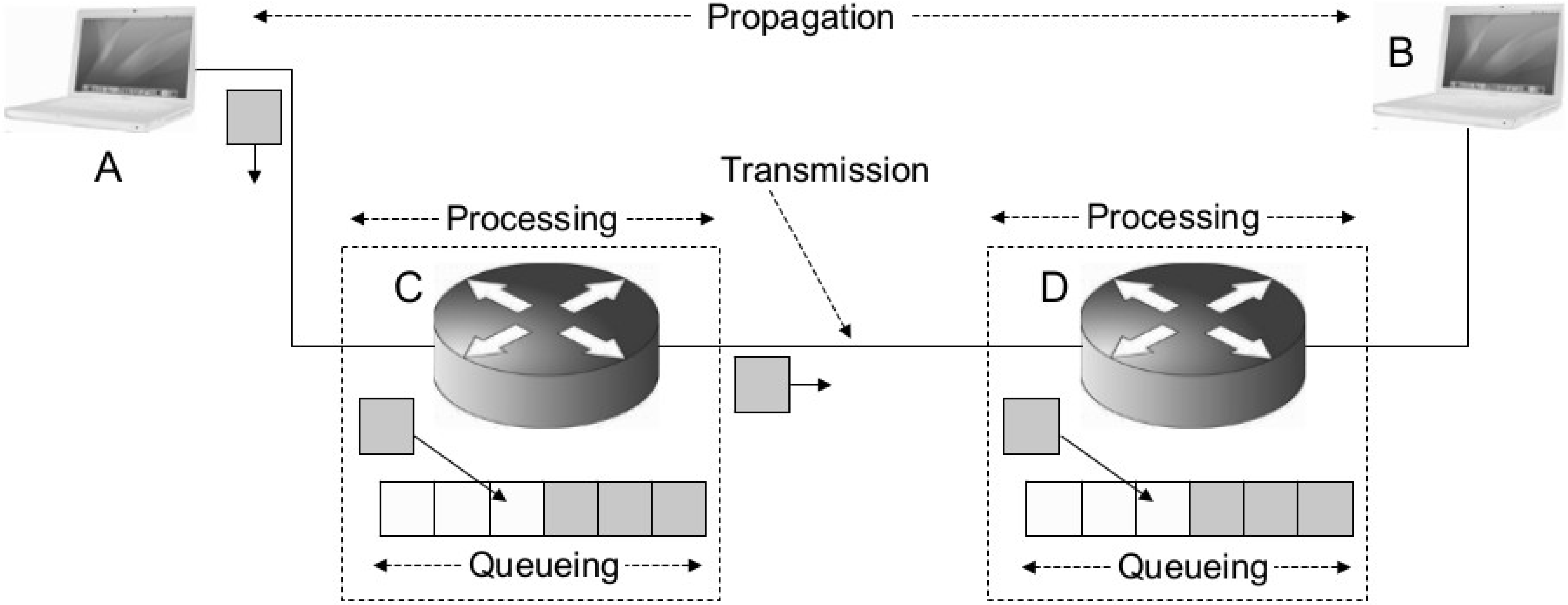
Assistant Professor
Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology
Room – 4A07
Email – adnan.eee@aust.edu
URL- http://adnan.quaium.com/aust/cse4295

Delay in Packet-Switched Networks



- Computer A sends a data packet to computer B. In its transfer through the network, the packet passes one or more switching computers (the routers C and D).
- One of the connections from router C along the transmission path leads to the next switching computer - the router D.
- This connection is upstream with its own queue at an output buffer. When a data packet reaches the switching computer C, the destination address is determined by the data in the packet header.

Delay in Packet-Switched Networks

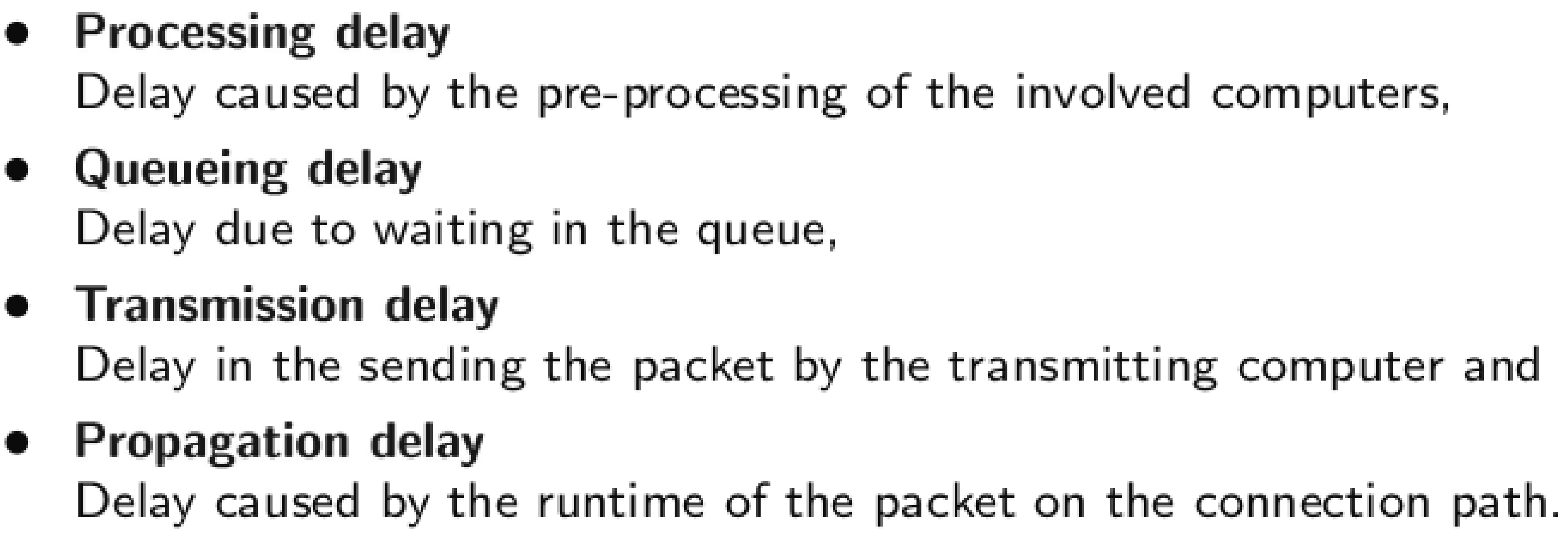


- In this way, the appropriate output connection is established. The data packet is now sent to router D via the output connection.
- It can only be forwarded via this output if the connection is not blocked by other data packets. These would already be at the output buffer ahead of the data packet to be forwarded.
- In this case, the data packet must line up in the output buffer queue.
- *Every computer the data packet passes causes delay.*

Delay in Packet-Switched Networks

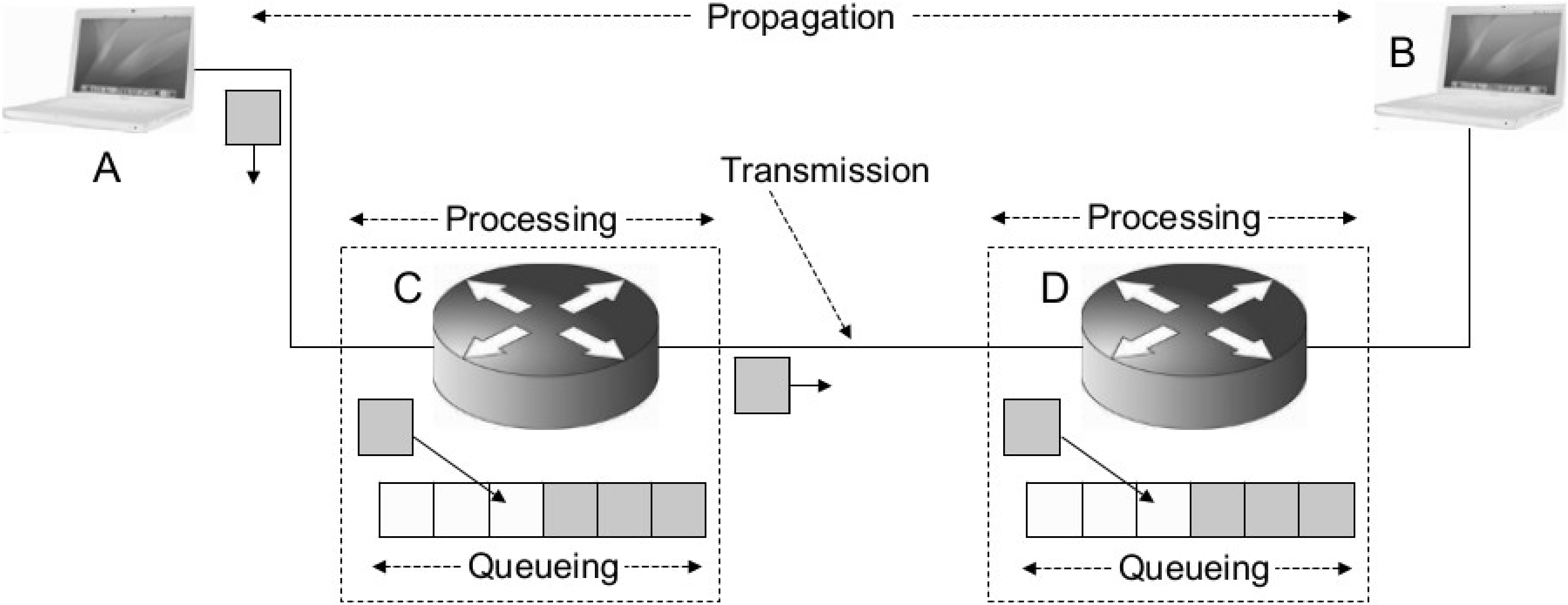
Every computer the data packet passes causes delay.

Reasons for this are:



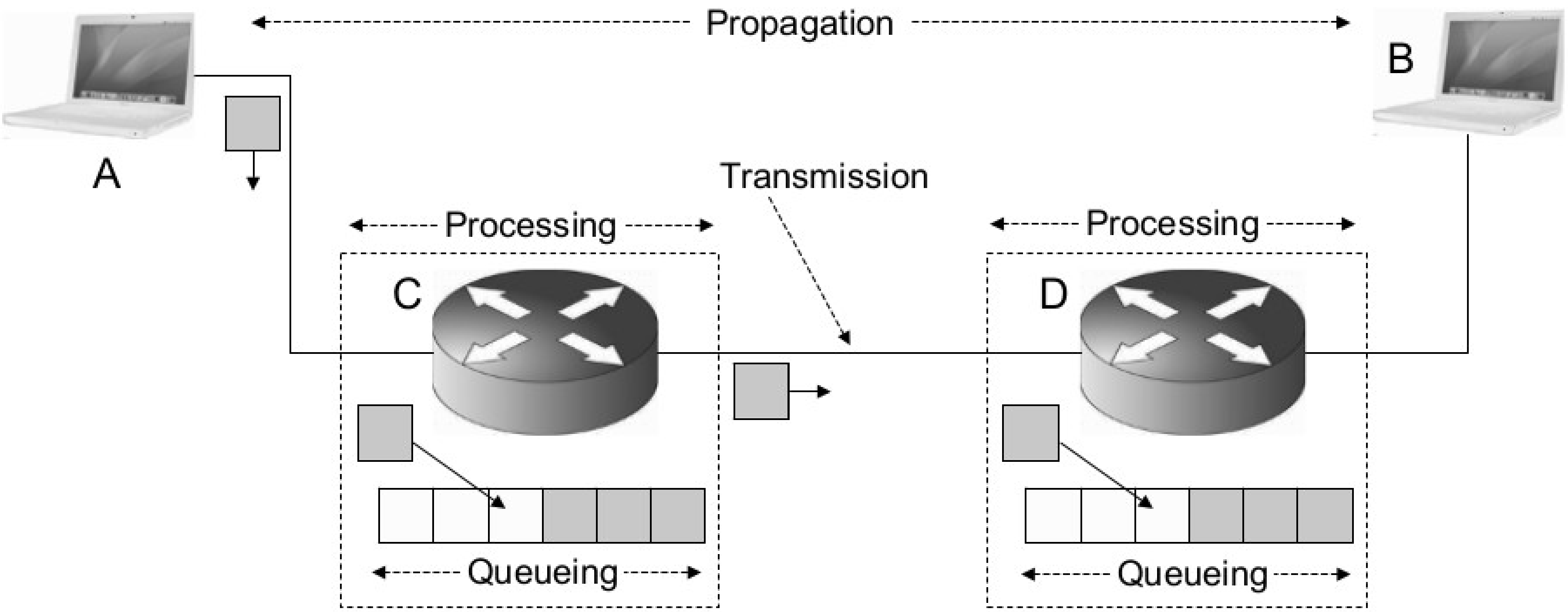
- **Processing delay**
  Delay caused by the pre-processing of the involved computers,
- **Queueing delay**
  Delay due to waiting in the queue,
- **Transmission delay**
  Delay in the sending the packet by the transmitting computer and
- **Propagation delay**
  Delay caused by the runtime of the packet on the connection path.

Processing Delay



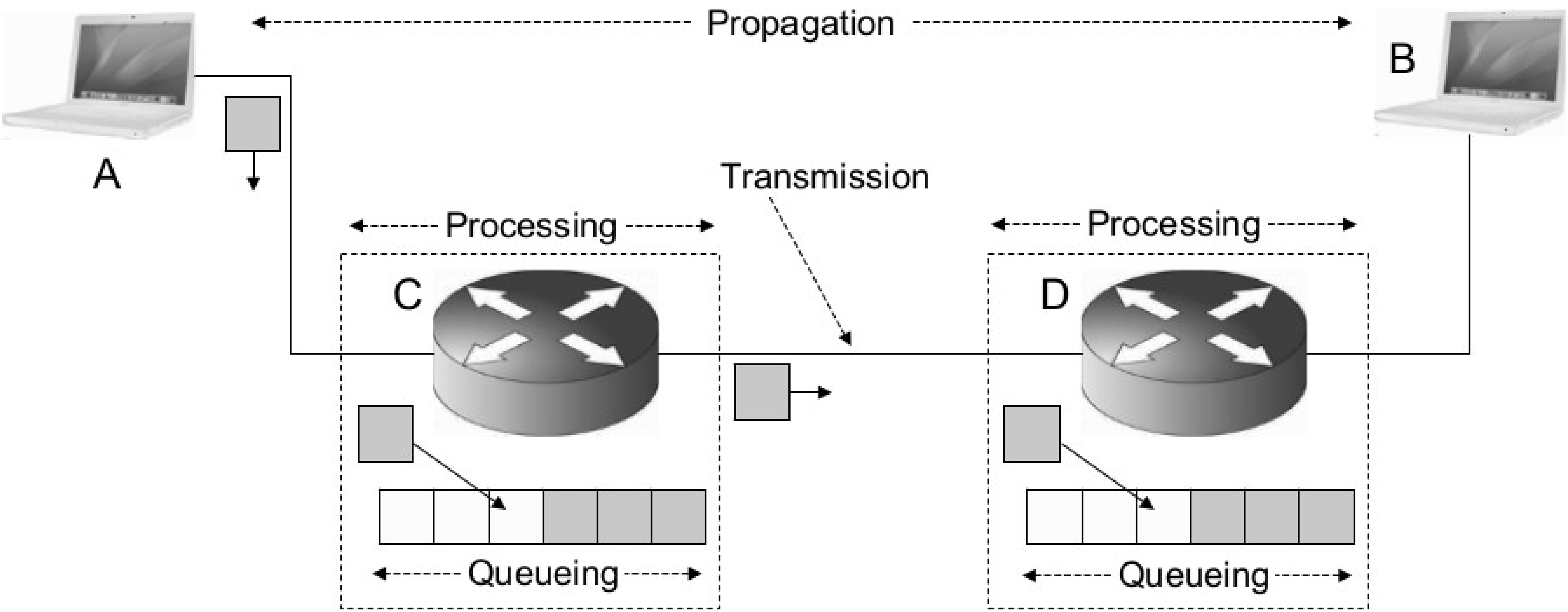
Processing delay
The processing delay d_{proc} is the time that the switching computer needs for pre-processing, i.e., to read the data packet header and to determine where the packet is to be sent. Calculated into this time is also the time necessary for error correction, if need be. In case bit errors have occurred in the transfer to the switching computer they can be recognized with the help of error detection methods and corrected in an appropriate manner. Today the processing delay at switching computers is on the order of microseconds and less. After this pre-processing, the switching computer directs the data packet to the output queue in the direction of router D.

Queueing Delay



Queueing delay
Packet transmission in the network proceeds on a strictly first-come-first-served basis. A packet can only first be transferred when all the packets that arrived before it at the output connection have been transferred. This means that while the packet is waiting in the queue at the output connection to router D for its transfer, it undergoes a delay d_{queue}. The delay is proportional to the number of data packets that are already in the queue. There is considerable variation in queue delay time. If the queue is empty, then the delay time is zero. The number of packets that an incoming data packet could encounter in the queue, depends on the intensity of network traffic. Practically-speaking, router queueing delay varies, ranging from the area of microsecond to millisecond.

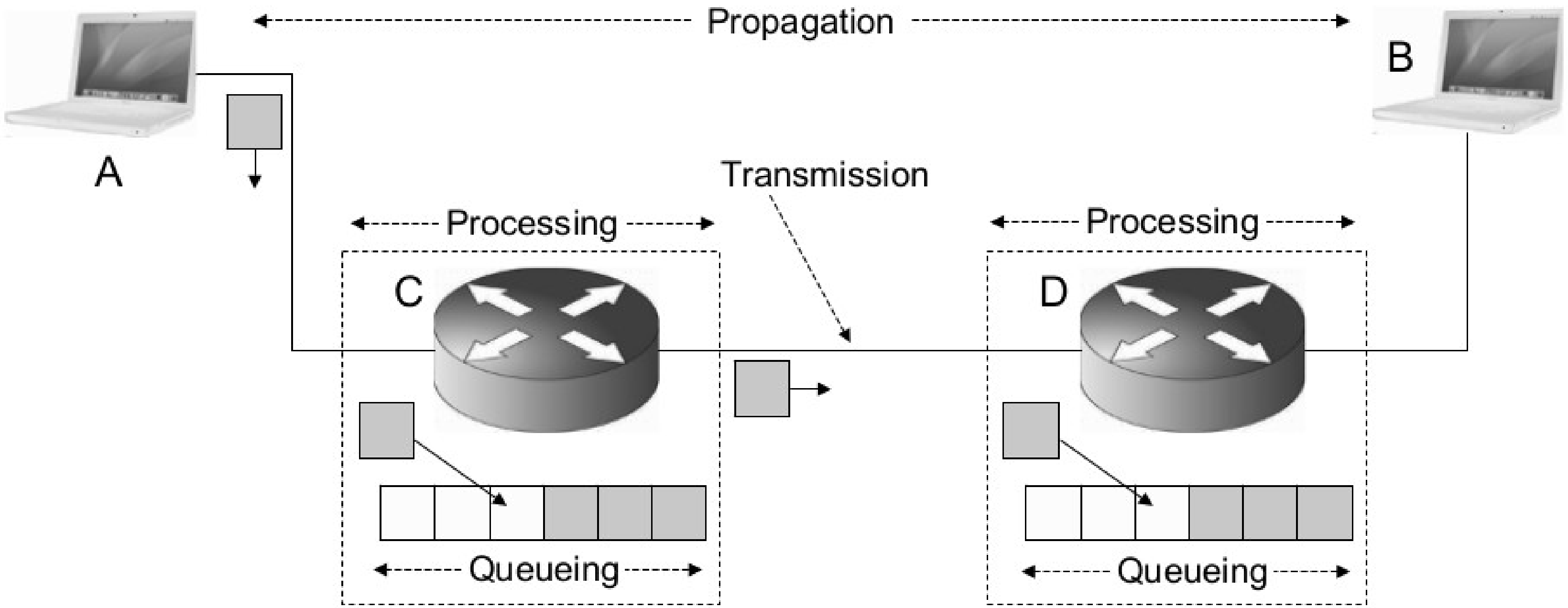
Transmission Delay



Transmission delay

If l designates the length of the data packet measured in bits and r the data transmission rate between router C and D measured in bit/second, then the transmission delay – also referred to as store-and-forward delay – is calculated simply as l/r. It defines the time required to deposit all of the packet's bit into the connection line. The transmission rate as such does not depend on the distance that the data packet has to bridge. It depends on how quickly the switching computer can transmit the data over the connection and also on the bandwidth of the connection. The transmission delay d_{trans} is in practical application in the microsecond range or less.

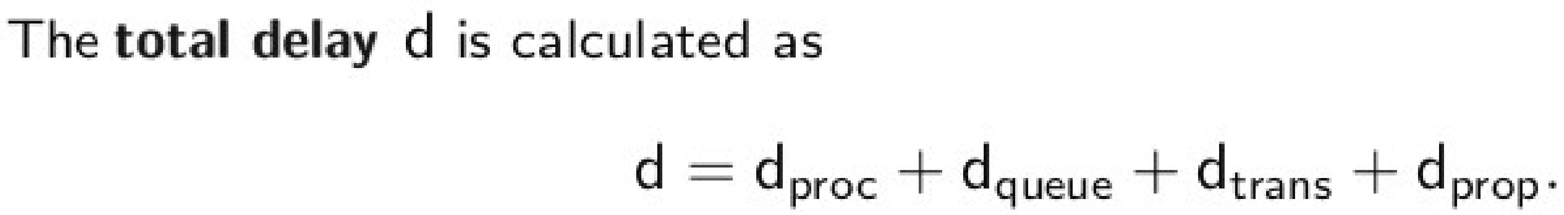
Propagation Delay



Propagation delay

If a data packet is sent on the connection between routers C and D it must follow the connection until router D is reached. The time between sending on the designated connection path and arrival at the receiver computer is referred to as propagation delay – d_{prop}. It is for the most part determined by the properties of the transmission medium (optical fiber, copper cable, radio waves, etc.) and at $2\text{-}3 \cdot 10^8$ m/s is close to the speed of light. The propagation delay is calculated from the distance b of the two switching computers divided by the propagation speed s, therefore b/s. When the last bit of the data packet has arrived at router D, it is stored temporarily with all of the previously sent bits of the data packet in router D. The entire process then repeats with router D as the starting point. In computer networks spanning large distances (WANs) the propagation delay d_{prop} is in the range of milliseconds.

Total Delay



The **total delay** d is calculated as

$$d = d_{\mathrm{proc}} + d_{\mathrm{queue}} + d_{\mathrm{trans}} + d_{\mathrm{prop}}.$$

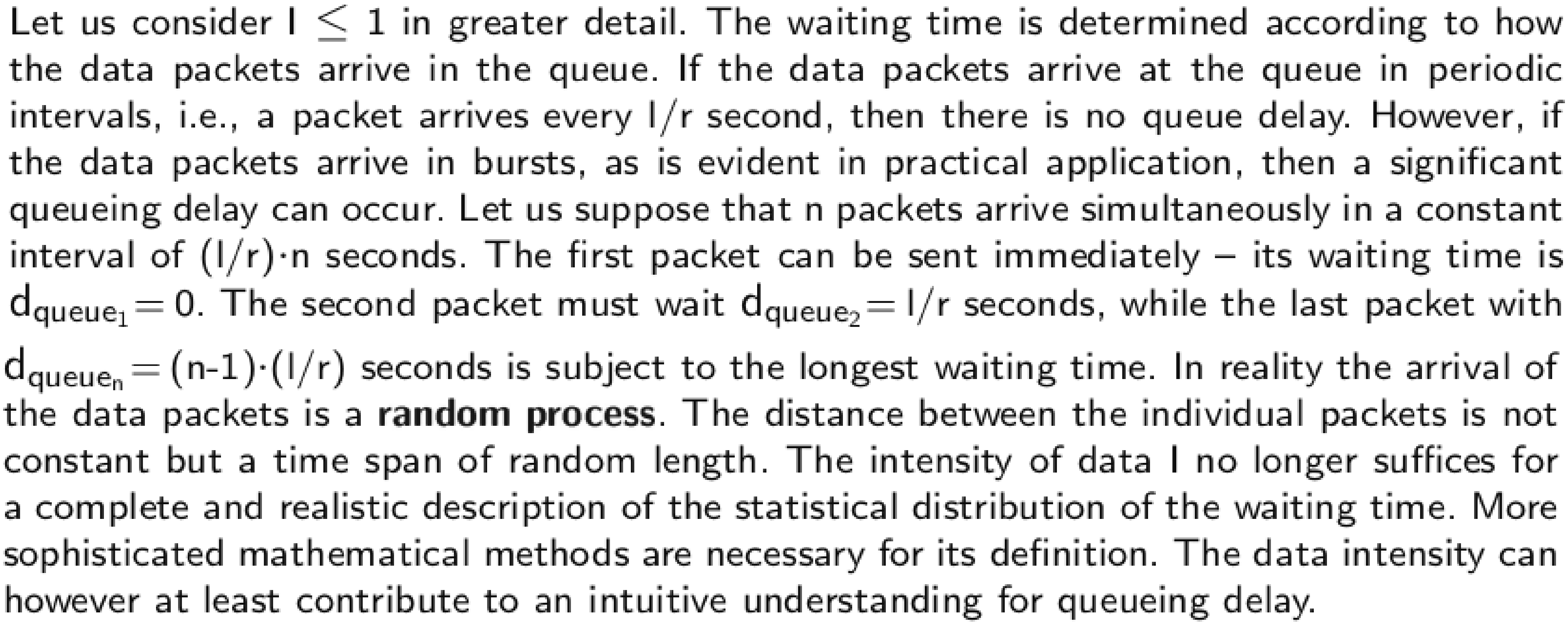
More about delay

In addition to the queueing delay d_{queue}, propagation delay d_{prop} plays the most important role. It extends from the microsecond range (e.g., if the routers are located in neighboring buildings) up to the area of several hundred milliseconds (e.g., if the routers are connected to each other via a satellite connection). There is also the transmission delay d_{trans}. It ranges from a negligible switching time (e.g., in the case of a 100 MBit Ethernet LAN) up to several hundred milliseconds (e.g., if data transmission is carried out via a slow 28 kbps modem).

More about delay

Special attention must be paid to **queueing delay**. Unlike other delay times, queueing delay depends on the respective network load and therefore can vary considerably from data packet to data packet. For example, if 10 data packets reach the queue at the same time, while the first packet can be transmitted without delay the tenth packet has to wait until all other nine packets have first been sent before its turn comes. **Statistical measures** are used to define the queueing delay. A definitive measure is the arrival time of the individual packets in the queue as well as how they are distributed. This means that the packets can arrive evenly distributed over time or cumulatively in so-called bursts . Let us look at the situation more closely: a stands for the average arrival rate of data packets at our queue, measured in packets per second; r is the transmission rate and l is the length of the data packet. The average arrival rate of the data is $l \cdot a$ bps. For the sake of simplicity, let us assume that the length of the queue is unlimited, i.e., no data packets can get lost. In this case, the ratio $I = \frac{l \cdot a}{r}$ the **intensity** of the data volume. If $I > 1$ then the average arrival rate of data is higher than the transmission rate of the connected line, i.e., the queue extends into infinity. It is therefore necessary that $I > 1$ always apply.

More about delay



Let us consider $I \leq 1$ in greater detail. The waiting time is determined according to how the data packets arrive in the queue. If the data packets arrive at the queue in periodic intervals, i.e., a packet arrives every l/r second, then there is no queue delay. However, if the data packets arrive in bursts, as is evident in practical application, then a significant queueing delay can occur. Let us suppose that n packets arrive simultaneously in a constant interval of $(l/r) \cdot n$ seconds. The first packet can be sent immediately – its waiting time is $d_{queue_1} = 0$. The second packet must wait $d_{queue_2} = l/r$ seconds, while the last packet with $d_{queue_n} = (n\text{-}1) \cdot (l/r)$ seconds is subject to the longest waiting time. In reality the arrival of the data packets is a **random process**. The distance between the individual packets is not constant but a time span of random length. The intensity of data I no longer suffices for a complete and realistic description of the statistical distribution of the waiting time. More sophisticated mathematical methods are necessary for its definition. The data intensity can however at least contribute to an intuitive understanding for queueing delay.

More about delay

- *If the intensity is near zero*, then the delay is negligible.
- *If the intensity is close to one*, time intervals occur in which the arrival rate exceeds the transmission rate and the data packets must wait in the queue.
- *If the intensity continues to be near one*, then the queueing delay grows rapidly. *Just a small percentage increase can lead to a tremendous rise in the queuing delay.*

Packet Loss

- In practical application this situation is naturally a different one as the *queueing capacity is always limited*. Thus, the queue cannot extend to infinity when data intensity approaches unity.
- If an incoming packet finds that the queue is filled, and there is no further memory available, then the switching computer cannot accept the data packet, i.e., *the packet is ignored and consequently lost*.
- This represents the *loss of a packet* for the end system: the packet was sent but never reached its destination.
- With a rise in data intensity, the number of packets that are lost increases.
- Therefore, the performance of a network computer is quantified by the probability of packet loss occurring, in addition to information concerning the average delay time.

Packet Loss

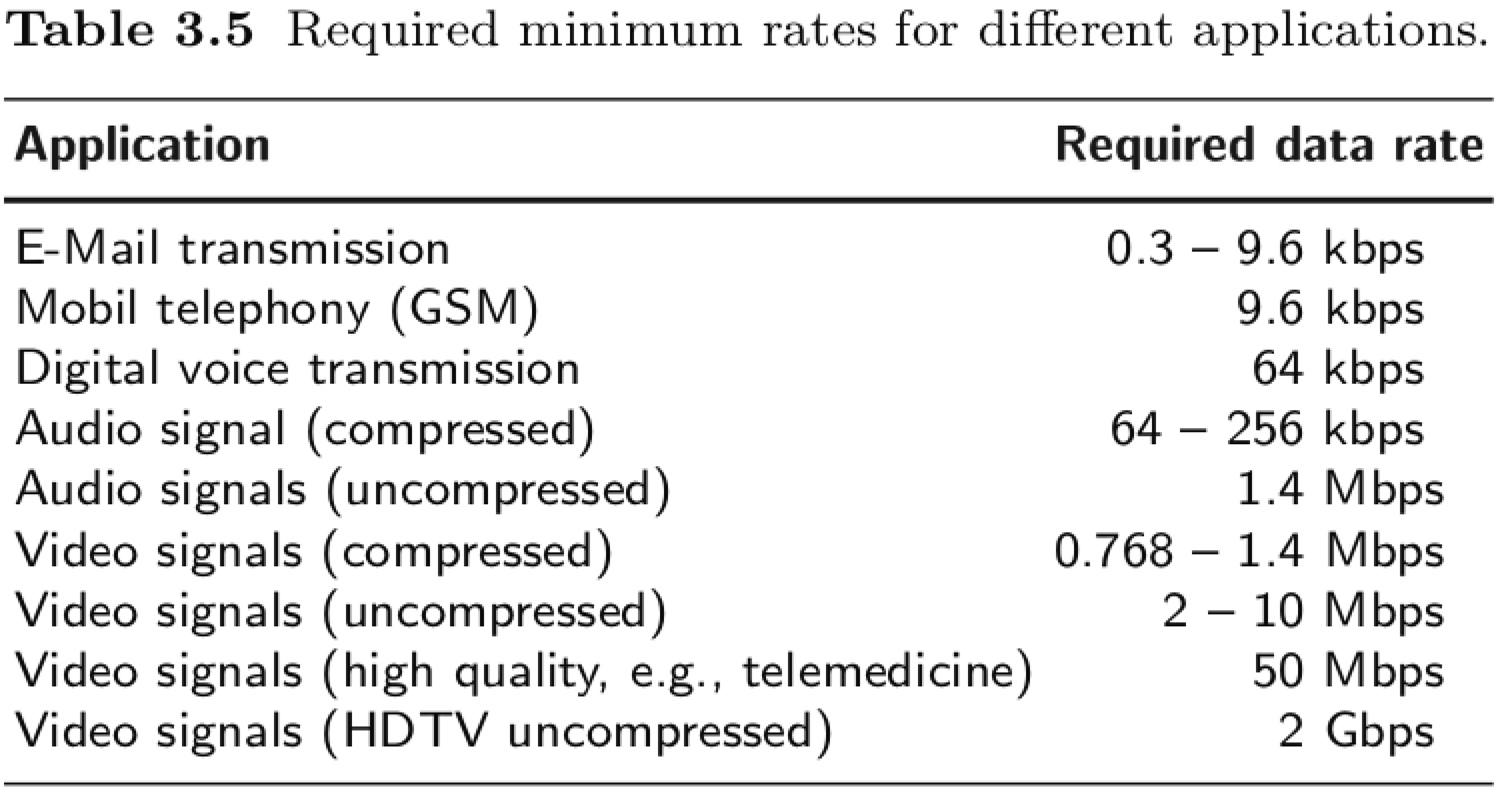


Table 3.5 Required minimum rates for different applications.

Application	**Required data rate**
E-Mail transmission	0.3 – 9.6 kbps
Mobil telephony (GSM)	9.6 kbps
Digital voice transmission	64 kbps
Audio signal (compressed)	64 – 256 kbps
Audio signals (uncompressed)	1.4 Mbps
Video signals (compressed)	0.768 – 1.4 Mbps
Video signals (uncompressed)	2 – 10 Mbps
Video signals (high quality, e.g., telemedicine)	50 Mbps
Video signals (HDTV uncompressed)	2 Gbps

Performance Characteristics of Communication Networks

- **Bandwidth :** Depending on the physical characteristics of a transmission medium, the bandwidth in telecommunication and communication engineering specifies a frequency range in which a signal, or data transmission, is even possible.

- **Data rate :** Depending on the signal coding used for data transmission, the available bandwidth of a transmission medium is limited and only a certain data transmission performance (data rate) can be achieved.

- **Throughput :** Refers to the actual amount of data transferred per time unit in a partial section of the network.

Performance Characteristics of Communication Networks

- **Runtime** : The runtime describes the time interval needed by a signal to travel across a transmission path. The runtime depends on the signal propagation speed of the respective transmission medium.

- Response Time :