

Lecture 4

Packet Jitter

1 Introduction

A crucial component of end-to-end delay is the varying queuing delays that a packet experiences in the network's routers. Because of these varying delays, the time from when a packet is generated at the source until it is received at the receiver can fluctuate from packet to packet (as shown in Figure 1-lecture 2). This phenomenon is called jitter. As an example, consider two consecutive packets in our VoIP application. The sender sends the second packet 20 msec after sending the first packet. But at the receiver, the spacing between these packets can become greater than 20 msec. To see this, suppose the first packet arrives at a nearly empty queue at a router, but just before the second packet arrives at the queue a large number of packets from other sources arrive at the same queue. Because the first packet experiences a small queuing delay and the second packet suffers a large queuing delay at this router, the first and second packets become spaced by more than 20 msec. The spacing between consecutive packets can also become less than 20 msec. To see this, again consider two consecutive packets. Suppose the first packet joins the end of a queue with a large number of packets, and the second packet arrives at the queue before this first packet is transmitted and before any packets from other sources arrive at the queue. In this case, our two packets find themselves one right after the other in the queue. If the time it takes to transmit a packet on the router's outbound link is less than 20 msec, then the spacing between first and second packets becomes less than 20 msec. If the receiver ignores the presence of jitter and plays out chunks as soon as they arrive, then the resulting audio quality can easily become unintelligible at the receiver. Fortunately, jitter can often be removed by using sequence numbers, timestamps, and a playout delay, as discussed below.

2 Removing Jitter at the Receiver for Audio

For our VoIP application, where packets are being generated periodically, the receiver should attempt to provide periodic playout of voice chunks in the presence of random network jitter. This is typically done by combining the following two mechanisms:

- Prepending each chunk with a timestamp. The sender stamps each chunk with the time at which the chunk was generated.
- Delaying playout of chunks at the receiver. As we saw in our earlier discussion of Figure 1-lecture 2, the playout delay of the received audio chunks must be long enough so that most of the packets are received before their scheduled playout times. This playout delay can either be fixed throughout the duration of the audio session or vary adaptively during the audio session lifetime.

We will examine two playback strategies: fixed playout delay and adaptive playout delay.

2.1 Fixed Playout Delay

With the fixed-delay strategy, the receiver attempts to play out each chunk exactly q msec after the chunk is generated. So if a chunk is timestamped at the sender at time t , the receiver plays out the chunk at time $t + q$, assuming the chunk has arrived by that time. Packets that arrive after their scheduled playout times are discarded and considered lost. What is a good choice for q ? VoIP can support delays up to about 400 msec, although a more satisfying conversational experience is achieved with smaller values of q . On the other hand, if q is made much smaller than 400 msec, then many packets may miss their scheduled playback times due to the network-induced packet jitter. Roughly speaking, if large variations in end-to-end delay are typical, it is preferable to use a large q ; on the other hand, if delay is small and variations in delay are also small, it is preferable to use a small q , perhaps less than 150 msec. The trade-off between the playback delay and packet loss is illustrated in Figure 1. The figure shows the times at which packets are generated and played out for a single talk spurt.

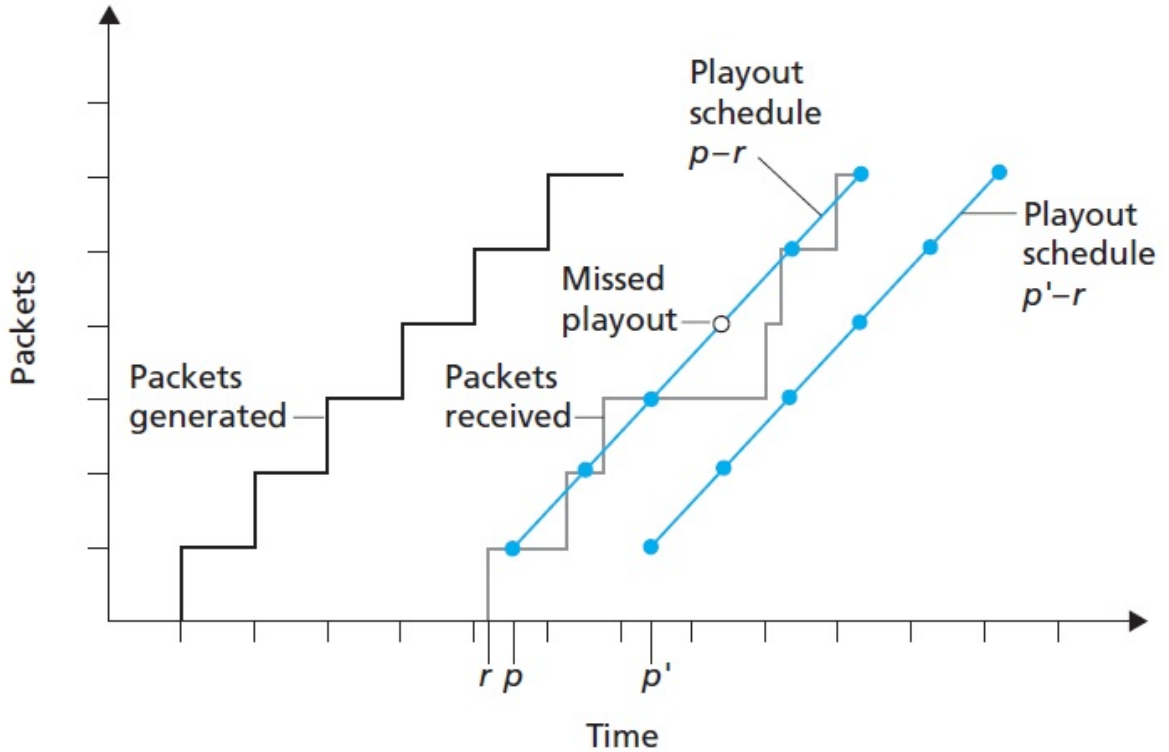


Figure 1: Packet loss for different fixed playout delays.

Two distinct initial playout delays are considered. As shown by the leftmost staircase, the sender generates packets at regular intervals—say, every 20 msec. The first packet in this talk spurt is received at time r . As shown in the figure, the arrivals of subsequent packets are not evenly spaced due to the network jitter. For the first playout schedule, the fixed initial playout delay is set to $p - r$. With this schedule, the fourth packet does not arrive by its scheduled playout time, and the receiver considers it lost. For the second playout schedule, the fixed initial playout delay is set to $p' - r$. For this schedule, all packets arrive before their scheduled playout times, and there is therefore no loss.

2.2 Adaptive Playout Delay

The previous example demonstrates an important delay-loss trade-off that arises when designing a playout strategy with fixed playout delays. By making the initial playout delay large, most packets will make their deadlines and there will therefore be negligible loss; however, for conversational services such as VoIP, long delays can become bothersome if not intolerable. Ideally, we would like the playout delay to be minimized subject to the constraint that the loss be below a few percent. The natural way to deal with this trade-off is to estimate the network delay and the variance of the network delay, and to adjust the playout delay accordingly at the beginning of each talk spurt. This adaptive adjustment of playout delays at the beginning of the talk spurts will cause the sender's silent periods to be compressed and elongated; however, compression and elongation of silence by a small amount is not noticeable in speech. We now describe a generic algorithm that the receiver can use to adaptively adjust its playout delays.

To this end, let:

t_i = the timestamp of the i th packet = the time the packet was generated by the sender

r_i = the time packet i is received by receiver

p_i = the time packet i is played at receiver

The end-to-end network delay of the i th packet is $r_i - t_i$. Due to network jitter, this delay will vary from packet to packet. Let d_i denote an estimate of the average network delay upon reception of the i th packet. This estimate is constructed from the timestamps as follows:

$$d_i = (1 - u)d_{i-1} + u(r_i - t_i)$$

where u is a fixed constant (for example, $u = 0.01$). Thus d_i is a smoothed average of the observed network delays $r_1 - t_1, \dots, r_i - t_i$. The estimate places more weight on the recently observed network delays than on the observed network delays of the distant past. Let v_i denote an estimate of the average deviation of the delay from the estimated average delay. This estimate is also constructed from the timestamps:

$$v_i = (1 - u) v_{i-1} + u \cdot |r_i - t_i - d_i|$$

The estimates d_i and v_i are calculated for every packet received, although they are used only to determine the playout point for the first packet in any talk spurt. Once having calculated these estimates, the receiver employs the following algorithm for the playout of packets. If packet i is the first packet of a talk spurt, its playout time, p_i , is computed as:

$$p_i = t_i + d_i + K v_i$$

where K is a positive constant (for example, $K = 4$). The purpose of the $K v_i$ term is to set the playout time far enough into the future so that only a small fraction of the arriving packets in the talk spurt will be lost due to late arrivals. The playout point for any subsequent packet in a talk spurt is computed as an offset from the point in time when the first packet in the talk spurt was played out. In particular, let:

$$q_i = p_i - t_i$$

be the length of time from when the first packet in the talk spurt is generated until it is played out. If packet j also belongs to this talk spurt, it is played out at time

$$p_j = t_j + q_i$$