This project on non-pre-emptive priority service will involve both analysis and simulation, in which the simulation is used to validate the analysis (and vice versa), to illustrate some characteristics of simulation estimates of statistics, and to obtain characteristics of priority service that are not amenable to simple analysis.

There is no specific format for the project report, but it should be complete and professional. The body of the report should be computer generated and printed (no hand-drawn figures or tables). However, the analytical calculations for parts 2a, 2c, 3a, and 3b may be done neatly by hand and included at the end of the report as an attachment. Simply refer to these calculations in the body of the report. However, the results of the analysis (plots, tables, etc.) must be computer-generated and printed as part of the body of the report. The report should include a cover page, an introduction, a brief description of the simulation model (part 1), and sections specifically for parts 2 and 3. The specific organization of the report for parts 2 and 3 is up to you, but should include presentation and discussion of results and conclusions.

1. Your first task is to build an Extend simulation model for the following situation. We want to model a single queue being fed by packets of up to three different priorities. Note that there is a priority queue (non-pre-emptive) built in to Extend’s Discrete Event library, and there are also Set Priority and Get Priority blocks built in. Your model must allow you to investigate the total delay statistics (queuing time plus service time) for each individual priority, as well as total delay statistics for all the packets (without dividing them into priority groups).

   In building your model, you must use the hierarchical block feature of Extend to keep your top-level simulation diagram relatively uncluttered. Also, your model should be focused on a packet transmission application, so that packet lengths and link bit rate should be specified separately.

2. We will begin by building on the results obtained in problem 2 of Assignment #10, in which you derived some results for a non-pre-emptive priority queuing system in which the service times of all P priorities are statistically identical. Here, we will be considering the following specific situation. There are two priorities, each with identical service time statistics. Let $E[B]=1$ and consider three cases of service time second moment: $E[B^2]$ equal to 1, 2, and 4. The system is designed for a high priority traffic intensity of 0.2 and a low priority traffic intensity of 0.5. We want to investigate what happens as one traffic intensity varies and especially what happens if the actual traffic exceeds the design values.

   a. Begin by computing total packet delay (queuing time plus service time) of each priority for the following scenario. Let the high priority arrival rate be 0.2 and vary the low priority traffic intensity from 0.5 to 0.9 in 0.1 increments. Note that the values of 0.8 and greater represent total traffic intensities of greater than 1.0. Be careful in your calculations for these cases (use the results of problem 3 of Assignment 10). Discuss what these results show about the behavior of priority service when (i) the second moment of the service time changes, and (ii) the low priority traffic exceeds the design value.

   b. Run simulations corresponding to the previous part. You need only simulate 3 values of $\rho_2$ for each value of $E[B^2]$: 0.5, 0.7, and 0.9 (9 simulations in all). Report the distribution and its parameters that you used for each value of $E[B^2]$. Each of the 9 simulations should
consist of 10 independent runs so that you could calculate confidence intervals for delay statistics (this will not be required, but some extra credit will be awarded if you do it and show the confidence intervals as error bars on result plots). Your "answers" will then be averages across the 10 runs. Use a run-in period of your choosing to mitigate initial transient effects in each run. Each run should be long enough so that each delay statistic for a given run is based on a minimum of 2,000 packets. That is, the minimum number of packets of a given priority in any given simulation run should be 2,000. Of course, some priorities will have many more than 2000 packets.

Summarize your results in a table (or set of plots) that compares the analytical results with the simulation results. For the simulation results, include the estimates of delay standard deviation (not available from our analysis) and the number of packets of each priority on which the simulation results are based, in addition to mean delay estimates.

Discuss the agreement between the results. In particular, do some results seem to agree better than others? If so, to what do you attribute this variation in agreement?

Be careful in your presentation and discussion of simulation results when the total traffic intensity is greater than 1.0. To illustrate the problems with this, conduct one additional simulation run with a traffic intensity greater than 1.0 and plot the individual packet delays (not the accumulated average delay) as a function of time.

c. Now use analysis (do not simulate) for the reverse situation, holding the arrival rate of priority 2 at 0.5 and investigating total delay statistics for priority 1 traffic intensities from 0.2 to 0.6. Again discuss what these results show about the behavior of priority service when (i) the second moment of the service time changes, and (ii) the high priority traffic exceeds the design value.

d. What are the fairness implications of your results? That is, imagine yourself in each of four situations: the four combinations of you being a high priority vs. low priority customer in the two situations of high priority traffic exceeding design value and low priority traffic exceeding design value. This part is important!

3. Now consider a more realistic application scenario. In the communications network for an international airline, three classes of messages are present as follows:

<table>
<thead>
<tr>
<th>Message Type</th>
<th>Priority</th>
<th>Message Length Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link Control</td>
<td>Highest (1)</td>
<td>Fixed at 128 bytes</td>
</tr>
<tr>
<td>Inquiry/Response</td>
<td>Next Highest (2)</td>
<td>Lognormal, Mean 1000 Bytes, S.D. 800 Bytes</td>
</tr>
<tr>
<td>Bulk Transfer</td>
<td>Lowest (3)</td>
<td>Uniform Between 0 and 10,000 Bytes</td>
</tr>
</tbody>
</table>

At one queue in this network, the link rate is 64 kb/s, and the total load (normalized traffic intensity) on the queue (link) is 0.7. Also, the relative arrival rates are $\lambda_1 = \lambda_2 + \lambda_3$ and $\lambda_2 = 4\lambda_3$.

a. Use analysis to find the following if the queue does not implement priorities.

i. Mean message delay (queuing time plus service time) per class.

ii. Mean message delay of all messages (traffic-weighted).
b. Repeat the previous part for non-pre-emptive priority service.

c. Simulate parts (a) and (b) above. The traffic-weighted mean should be directly measured from the simulation, not calculated from the per-class means. Report means (check against analysis) and standard deviations (not available from analysis) of message delays.

d. Briefly discuss the results obtained in this application section.