General Description of the problem:
In this assignment, we will approximate a client sending requests to a server. Each request is a pair of 32x32 grayscale images (8-bit per pixel). The server then equalizes the image histogram using the algorithm from homework 1 (modified for smaller images), and returns a result.

In order to simplify things, we are not going to build an actual client-server application, and will not use any networking. Instead, we will “emulate” the client within our application as described in this pseudo-code:

```
for i in [0 .. NREQUESTS - 1] {
    checkForCompletedRequests();
    // emulate a certain request rate from the client
    waitRandomTime(load);
    processImageOnGPU(&images_in[i], &images_out[i]);
}
waitForRemainingRequestsToComplete();
```

We will implement the server functionality (`processImageOnGPU` in the pseudo-code) twice: once using CUDA streams, and another using producer consumer queues between CPU and GPU.