```
period = 20 seconds
lower_threshold = 2 seconds
upper_threshold = 5 seconds

while True:
      start = get_time()
      sleep(period)
      if active:
            f = open(nginx_log_path)
            times = []

            for record in f:
                  if record.time > start:
                        times.push(record.response_time)
            average_time = average(times)
            if average_time > upper_threshold:
                  scale_up()
            else if average_time < lower_threshold:
                  scale_down()

// nginx updates log files for each requests and record their response
time, please refer to the project/nginx/nginx.conf for log_format
// below is a pseudo-behavior for nginx updating the log file
while True:
      handle_requests()
      record = measure_response_time()
      write_to_log_file(record, nginx_log_path)
```

==========

Period is set to 20 seconds to avoid the ping-pong effect.

Lower_threshold is set to 2 seconds as that is the observed average response time from a single instance under relatively low workload.

Upper_threshold is set to 5 seconds because we observed a linear relationship between response time and workload for a single instance, so with 5 seconds of response, on average, the instance is overloaded with more than 2 concurrent requests, therefore we scale up.