Wait Time in FIFO Queues

Kyriafinis Vasilis, 9797

March 11, 2022

Introduction

The objective of this project is to measure and optimize the wait time of processes ready to execute using a FIFO queue for scheduling. The setup is simple, p threads are used as producers inserting "work" in the queue and q threads representing consumers execute the work previously inserted in the queue.

Three parameters can be altered to find the optimum number of consuming threads.

1. Number of producers (p): The number of work producing threads.

2. Number of consumers (q): The number of work consuming threads.

3. Queue size: The maximum number of work units that can be pending at any given moment.

Testing

The computer used to run the tests has a *Ryzen* 7 5700*G* processor with 8 cores and 16 threads running at 3.8 *GHz*. The goal for this experiment is to always keep the consuming threads busy. To achieve this for the initial test the number of consumers and producers is the same. The consuming threads are slower than the producing threads by design. This forces the queue to fill up.

Queue size

To observe the effect of the queue size, one producer and one consumer are created. Once the queue is full no more work units are produced until the consuming thread executes a previously inserted unit. Because the consuming thread is slower than the producing thread the bigger the queue size the bigger the wait time becomes.
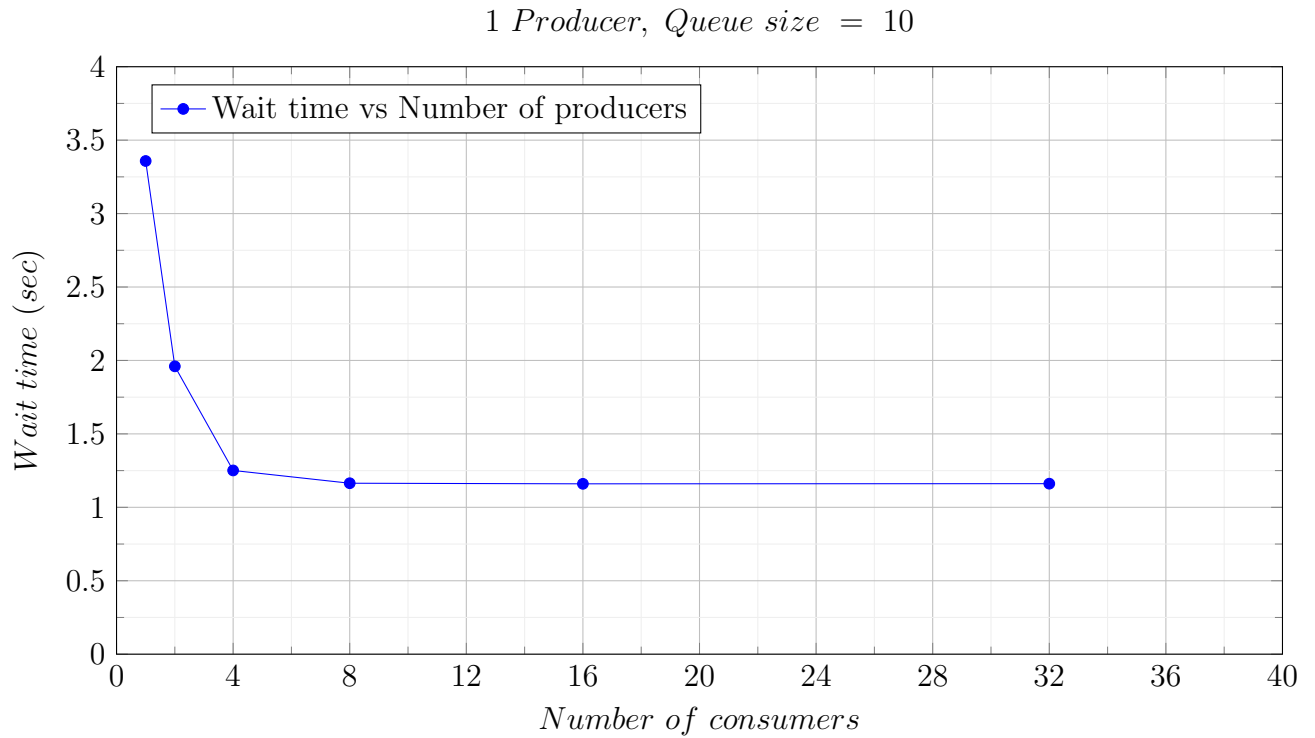

1 *Producer*, 1 *Consumer*
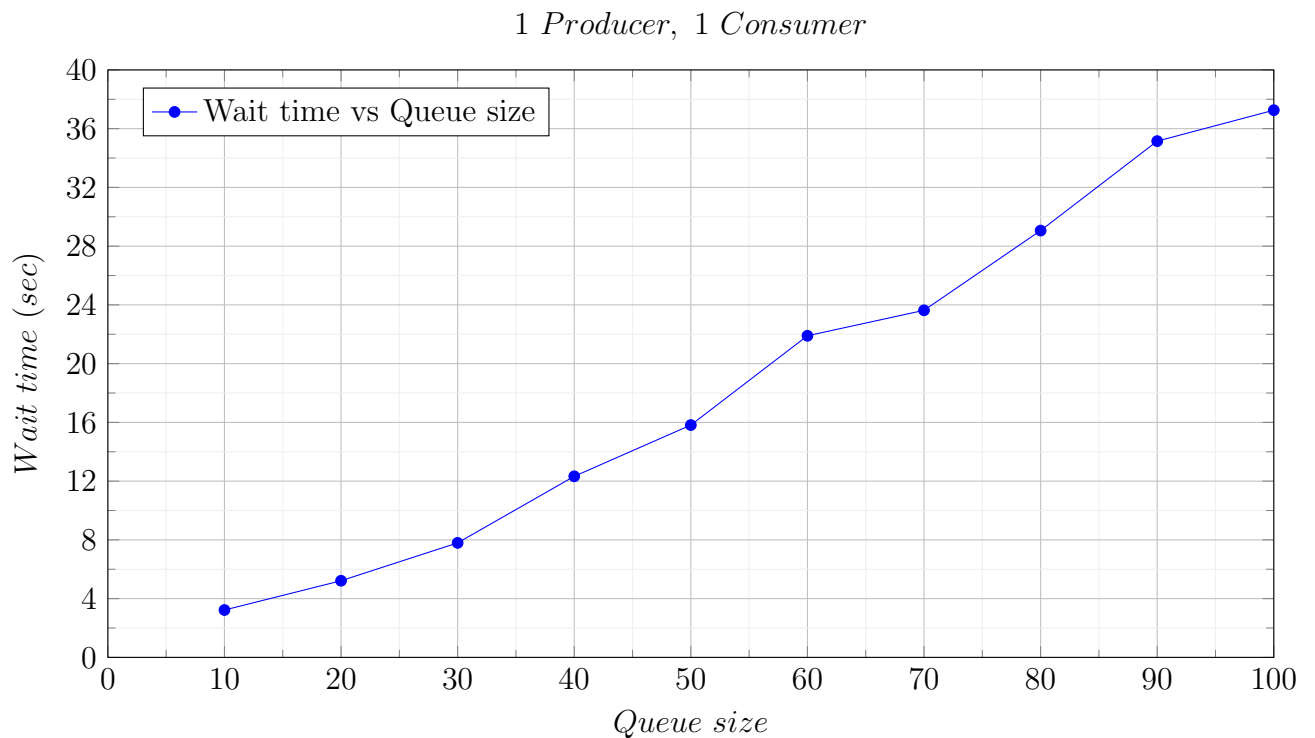
Producers Number

Given a queue size, the number of producers must be a fraction of the consumer number. This is because if the producers are equal or more than the consumers the queue fills up and the producers stop creating more work. For this reason, initially, only one producer is created. After defining the optimal number of consumers the $\frac{producers}{consumers}$ fraction is determined. If this fraction is kept steady the wait time is relatively stable.

Consumers Number

Finally starting from one consumer and increasing the number until the wait time stops decreasing the optimal number of consumer threads is determined. This number is always relative to the producers number.

$1\ Producer,\ Queue\ size\ =\ 10$

The graph above suggests that for every producer 8 consumers are required to keep an optimum configuration. If more consumers are added the queue will be empty for the most time and consuming threads will wait for producing threads to create new work. In contrast, if more producing threads are spawned the queue will be full for the most time and the waiting time will increase.