### M/M/S排队模型模拟

## 1 实验背景

排队问题在系统随机聚散现象和随机服务系统的工作过程中都有极为重要的地位，在计算机网络中，排队问题主要出现在通信传输的过程中，对于网络传输中多队列但服务台模型，不同队列间的公平性尤为重要。

IEEE802.11的MAC协议定义了分布式协调功能（DCF，Distributed Coordination Function）和点协调功能（PCF，Point Coordination Function）两种接入机制，本实验通过模拟一个带有中央控制的多队列单服务台排队模型，对PCF机制下的控制方法进行研究，尽可能地保证多队列间的公平性。

## 2 实验目的

1. 进行单队列模拟实验，其中包的长度服从指数分布，包的达到情况服从泊松过程，探究队列长度的分布情况与包等待时间的分布情况。
2. 进行三队列模拟实验，其中包的长度服从指数分布，包的达到情况服从泊松过程，设计带有中央控制的三队列调度策略，通过比较三个队列的等待时间分布、吞吐量、队列长度分布等方案，探究调度的公平性，通过改进调度策略使其尽可能地公平。

## 3 实验过程及内容

### 3.1单队列模拟实验

单队列模拟实验中，到达的包的服务顺序为先来先服务，具体模拟的分析与结果如下：

**3.1.1包的到达模式**

有无限个数的包可以到达，模拟仿真时通过设置一个较大的发包数n来模拟这种情况，包的到达服从参数为的泊松过程，即时间t内有一个包到达的概率为：



经过数学变换可得，包的到达间隔为:

其中p为一个随机的到达概率。

**3.1.2包的服务模式**

按照先来先服务的顺序对包服务，每个包的长度服从参数为的指数分布，则每个包的长度为：



其中p为一个随机的到达概率。

设服务台以固定的速率对到达的包进行服务，在该实验中我们将设置为1，则每个包的服务时间t为：

**3.1.3实现方法**

模拟的方法主要分为以下几步：

* 根据上述包的到达模式与服务模式，产生单队列中n个包的到达时间与长度，存储在一个数组模拟的队列中。
* 维护系统当前时间变量，主要处理过程为一个循环体，伪代码如下：

**全局变量定义:**

常量V:服务台的服务速率

curTime:系统当前时间

queue:存储包的队列

headPack:队首包，属性arriveTime记录了包的到达时间,属性length记录了包的长度，属性serveTime记录了包开始服务时间

**模拟算法:**

while(queue中有包剩余){

headPack=queue.head

//队列中包在当前时间未到达时移动当前时间至下一个包到达

if(headPack.arriveTime>curTime){

curTime=headPack.arriveTime

//记录包的开始服务时间并移动当前时间至headPack完成服务

headPack.serveTime=curTime

curTime+=headPack.length/v

//队列出队并更新headPack

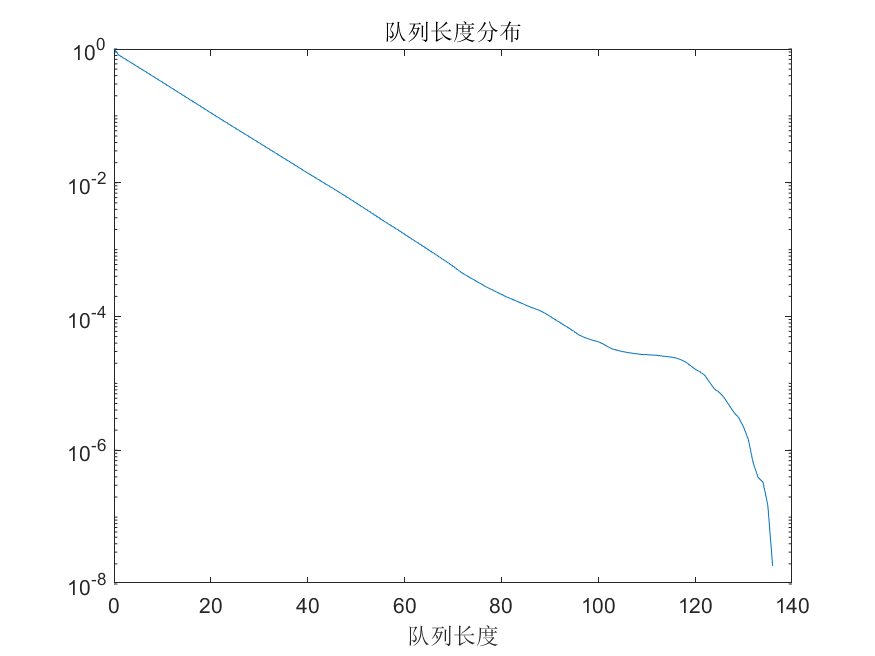
queue.pop()

}

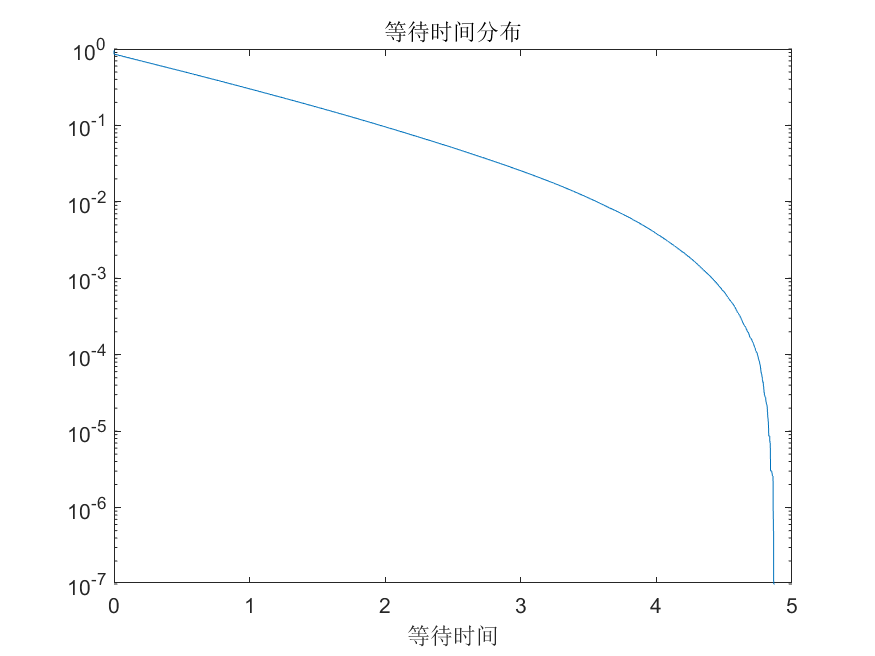
* 根据记录的包的长度，计算模拟过程中队列的等待时间分布、队列长度分布、平均等待时间，平均队列长度以及吞吐量，服务台的占用率等指标，计算过程略去。

**3.1.4模拟仿真与结果**

1. 设置=9,=10，发包的数量n=，绘制在队列不饱和情况下的队列长度分布与等待时间分布图像如下：



=9,=10的M/M/1队列长度分布图



=9,=10的M/M/1等待时间分布图

模拟得到的其他指标如下所示：

|  |  |  |
| --- | --- | --- |
| 平均等待时间 | 平均队列长度 | 服务台占用率 |
| 0.90903 | 8.18183 | 89.97% |

在的情况下（即队列不饱和），M/M/1队列可通过数学推导计算出以上三项指标，公式如下：

平均等待时间 = 

平均队列长度 = 

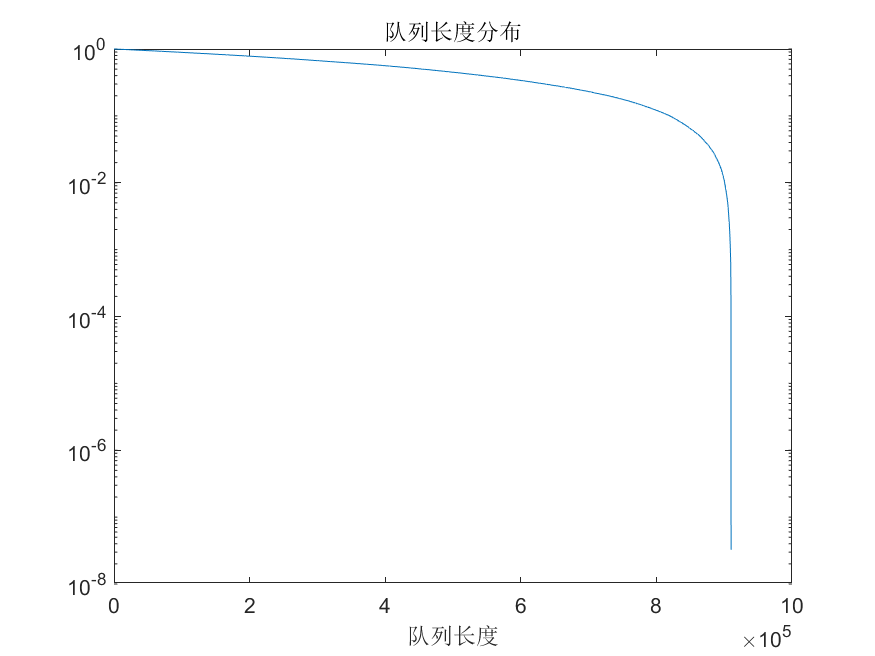
服务台占用率 = 

计算可得三项指标理论指如下：

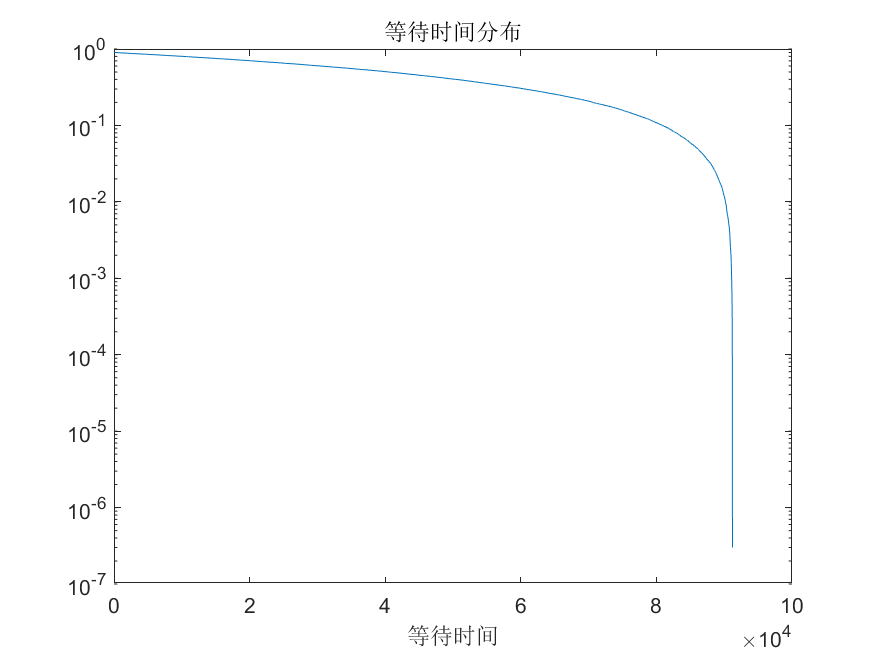
|  |  |  |
| --- | --- | --- |
| 平均等待时间 | 平均队列长度 | 服务台占用率 |
| 0.90000 | 8.10000 | 90.00% |

模拟结果与理论计算相符合。

1. 设置=11,=10，发包的数量n=，绘制在队列饱和情况下的队列长度分布与等待时间分布图像如下：



=11,=10的M/M/1队列长度分布图



=11,=10的M/M/1等待时间分布图

模拟得到的其他指标如下所示：

|  |  |  |
| --- | --- | --- |
| 平均等待时间 | 平均队列长度 | 服务台占用率 |
| 45516.09661 | 455001.85179 | 100% |

通过模拟可以看出，在队列饱和的情况下，由于服务台在多数时候无法在下一个包到来前处理完当前服务的包，因此出现了队列无限增长的情况。

### 3.2多队列模拟实验

对三队列单服务台模型进行模拟，采用赤字轮询的方式进行调度，探究多队列下统一调度方案的公平性。

**3.2.1 包的到达模式**

每个队列都有无限个包到达，模拟仿真时通过设置到达的包数为一个较大值来模拟这种情况，每个队列中包的到达是相互独立的过程，均服从泊松过程，泊松过程的描述详见3.1.1。

**3.2.1包的服务模式**

按照赤字轮询的方式决定下一个服务的队列，在同一队列中包按照先来先服务的顺序进行服务。

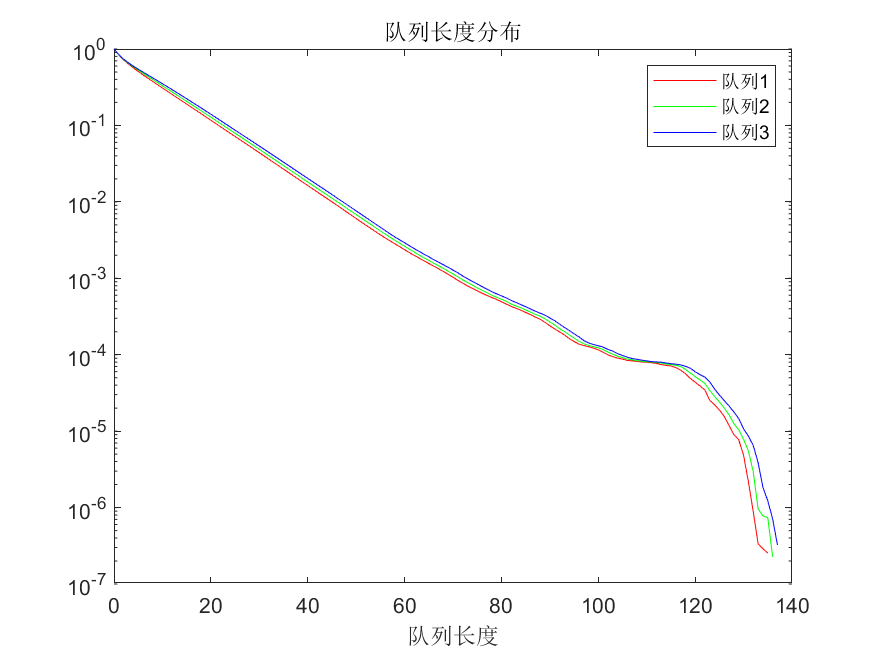
传统的时间片轮转调度方案会为每个队列分配相同的时间片大小，当一个队列用尽一个时间片的时间后，即使当前有包处在处理的过程中，服务台依然会终止对其的服务，转而去服务下一个队列，这显然不符合传输的时机情况，包的服务过程被中断是不被允许的。

赤字轮询是对传统的时间片轮转调度方案的改进，在赤字轮询调度方案中，将时间片的概念替换为轮询一次可服务的包长总和，同时加入赤字的概念，对包长超过该队列剩余可服务包长的情况进行处理，当上述情况出现时，服务台将不会再对该队列进行服务，而是将该队列剩余的可处理包长大小大小记录赤字，用于扩充下一次轮询到该队列时可处理的包长，即每次队列开始服务的可处理包长大小为上一次的赤字加上为每个队列固定分配的可处理包长大小。这种调度方式在较长时间来看，不同队列间的吞吐量指标上是公平的。值得注意的是，由于在实际情况中服务台的服务速率可能变化，因此对包长的赤字相比对时间记录赤字更加公平。

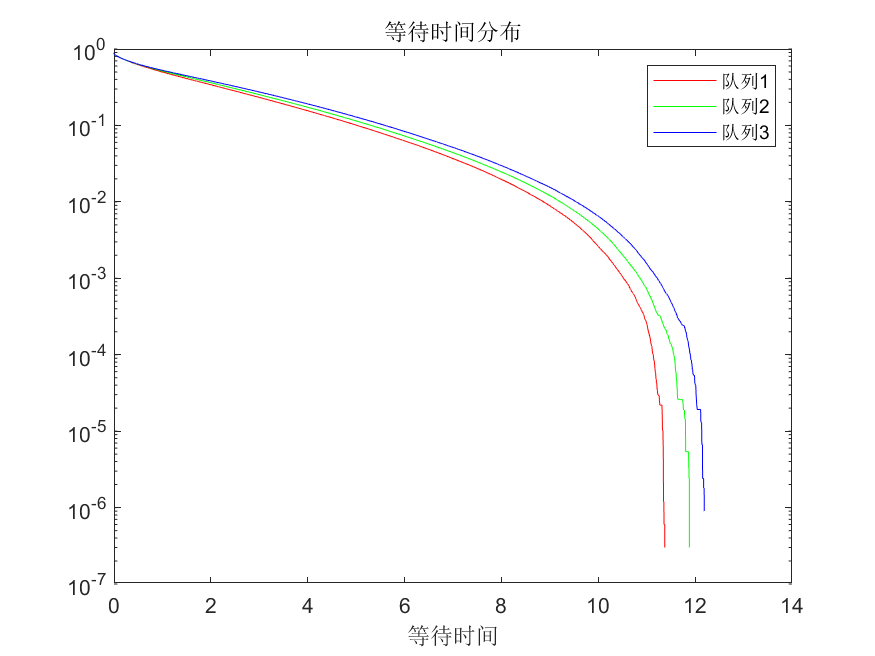
包的长度与服务时间的计算详见3.1.2。

**3.2.3模拟仿真与结果**

(1)设置三队列，，轮询一次可处理包长大小设置为0.4，总的发包的数量n=，此时三队列队列长度分布与等待时间分布如下所示：



,队列长度分布图



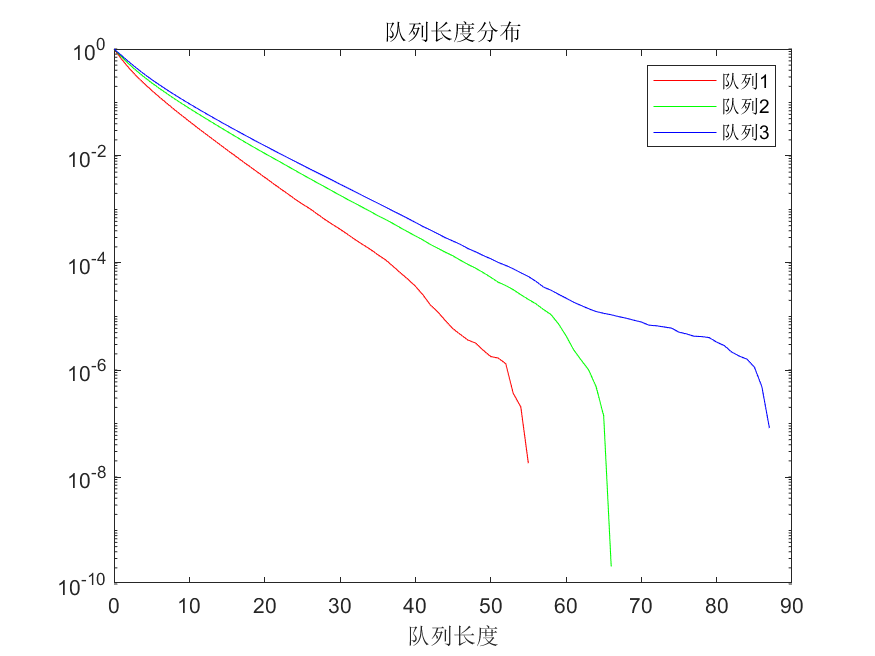
,等待时间分布图

模拟仿真得到的其他指标如下所示，服务器利用率（占用率）为89.68%：

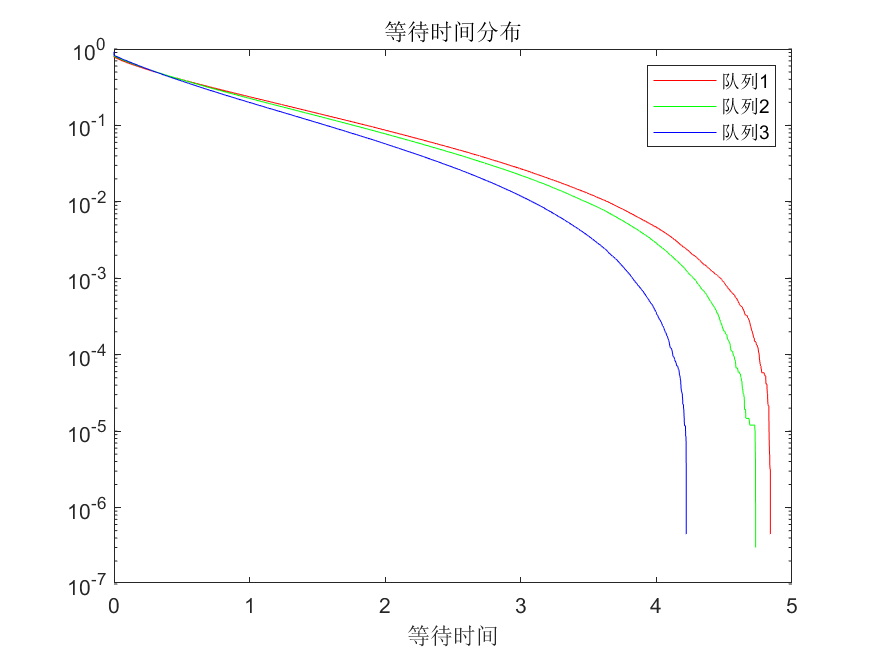
|  |  |  |  |
| --- | --- | --- | --- |
|  | 平均等待时间 | 平均队列长度 | 吞吐量 |
| 队列1() | 0.87 | 8.18 | 333469.07 |
| 队列2() | 0.93 | 8.60 | 333468.88 |
| 队列3() | 0.96 | 9.08 | 333469.07 |

可以看出，在设置三队列的与值相同时，三队列的队列长度分布、等待时间分布与各项指标都是相近的，且与单队列模拟设置=9,=10时的情况十分类似，也在一定程度上验证了模拟仿真结果的正确性与调度方法的公平性。

(2)设置三队列的，，轮询一次可处理包长大小设置为0.4，总的发包的数量n=，此时的三队列队列长度分布与等待时间分布如下所示：



,队列长度分布图



,等待时间分布图

模拟仿真得到的其他指标如下所示，服务器利用率（占用率）为90.05%：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 平均等待时间 | 平均队列长度 | 吞吐量 |
| 队列1() | 0.24 | 2.35 | 222320.12 |
| 队列2() | 0.32 | 3.14 | 333384.41 |
| 队列3() | 0.35 | 3.58 | 444663.82 |

通过上述数据可以发现，三队列的吞吐量大致以2:3:4成比例，即为每个队列的之比，似乎三队列在吞吐量上并不公平。实际是因为队列并不处于饱和状态，对于三条队列，总共只有等于队列对应吞吐量长度的包到达，几乎没有包停留在队列中得不到服务，因此吞吐量将取决于模拟仿真过程中美国队列发包的数量与包的长度，为了验证赤字轮询在吞吐量上是公平的，记录仿真结束时刻三队列的剩余包长赤字大小如下：

|  |  |
| --- | --- |
|  | 剩余包长赤字大小 |
| 队列1 | 330578.27 |
| 队列2 | 219513.98 |
| 队列3 | 108234.57 |

可以看出，三队列吞吐量与剩余可包长赤字大小之和是相近的，假设此时网络突然变得繁忙，在一段时间内，三队列得到服务的优先级从大到小为队列1、队列2、队列3，直到三队列的剩余包长赤字大小相近，即吞吐量相近。因此可以得出结论，从较长时间来看，赤字轮询调度是一种在吞吐量上公平的调度方法。

此外，注意到在等待时间分布图中三条队列曲线的相对位置顺序与队列值大小顺序相反，队列1()对应的红色曲线在等待时间分布图中位于三条曲线的最上方，以横坐标为3的红色曲线上的点为例，从图中可以看出，p(队列1中的包等待时间>3)>p(队列2中的包等待时间>3)p(队列2中的包等待时间>3)，与三个队列包到达频率的大小恰好相反，我们认为，由于每条队列均处于不饱和的状态，每个队列一次轮询中处理的包长小于分配的可处理包长，因此每个队列在一轮轮询的过程中等待的时间取决于其他队列到达包的服务时间之和，在理想情况下进行分析，假设每个队列的包到达间隔与包长均为定值，则服务台一轮轮询时间也为定值，设为t，且假设包到达间隔小于服务台一轮轮询的时间，一轮轮询中到达的包均能在这一轮中处理完，由于三队列值相等，设每个包的长度均为，每个队列一次轮序中等待时间如下：

一次轮询中队列1的等待时间

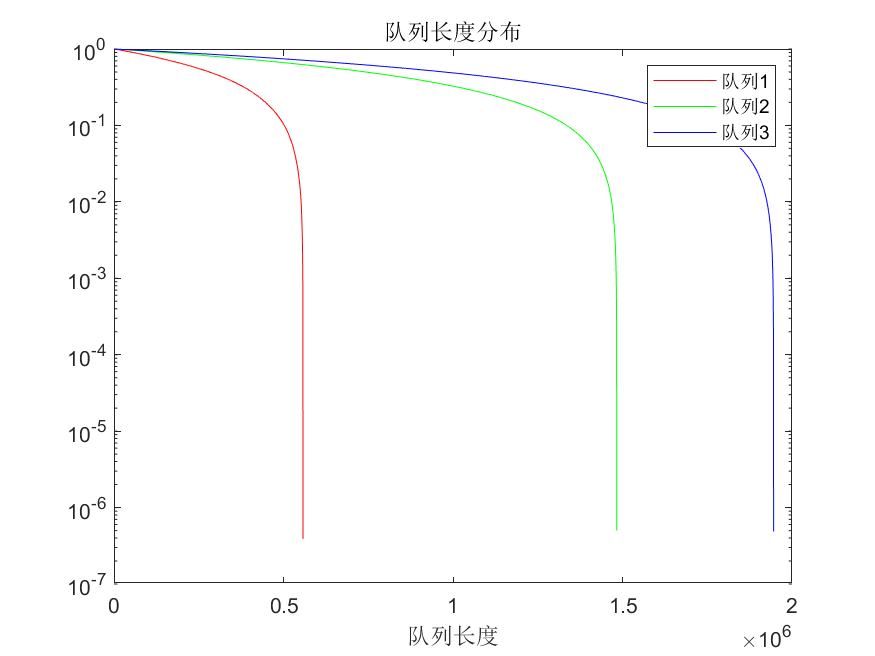
一次轮询中队列2的等待时间

一次轮询中队列3的等待时间

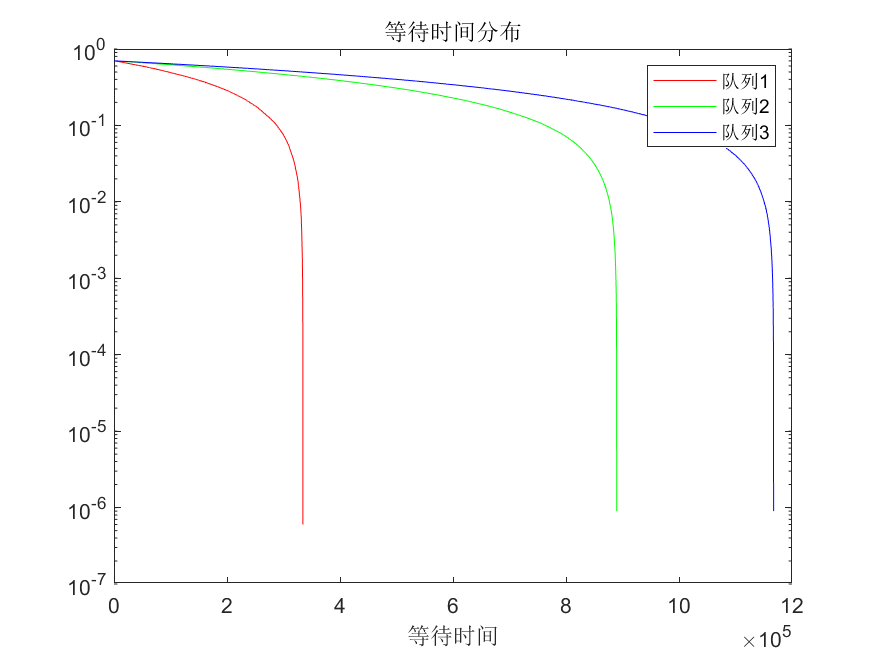
由于队列1的包到达频率最低，因此出现了队列1的包有更高的概率等待较长的时间的情况。

根据上述分析，我们猜想在队列饱和情况下三队列吞吐量应当相等且在等待时间分布图中的相对位置应与队列值大小一致。

1. 为了验证在(2)中的猜想正确，我们在三队列均达到饱和的状态下模拟调度过程，设置三队列的，，轮询一次可处理包长大小设置为0.4，总的发包的数量n=，此时的三队列队列长度分布与等待时间分布如下所示：



,队列长度分布图



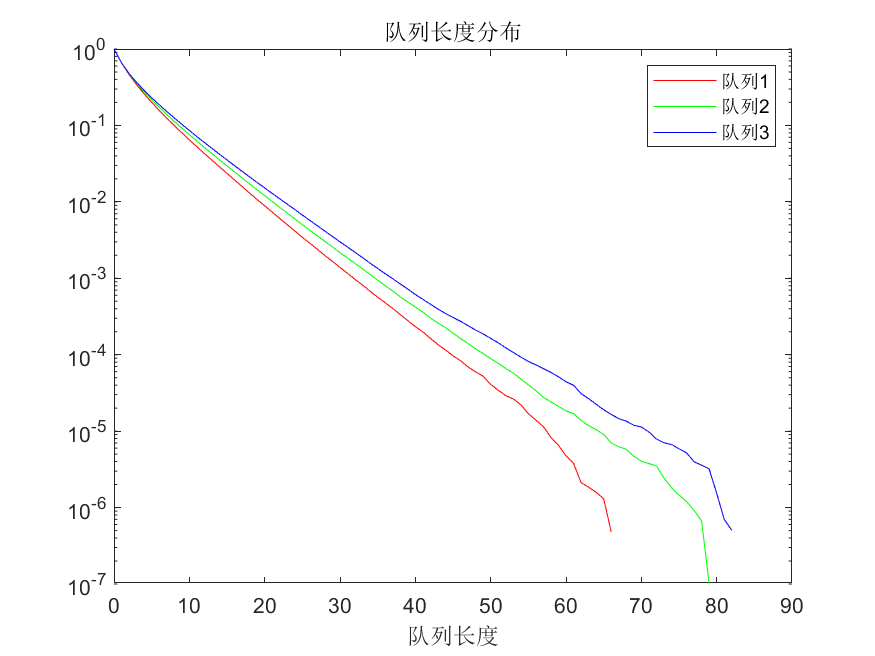
,等待时间分布图

模拟仿真得到的其他指标如下所示，服务台利用率（占用率）为100%：

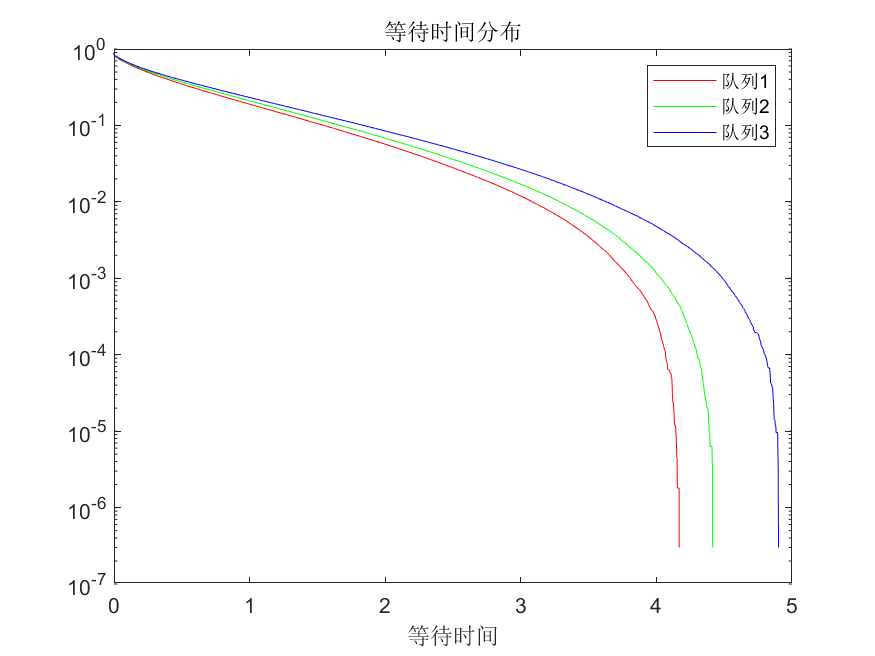
|  |  |  |  |
| --- | --- | --- | --- |
|  | 平均等待时间 | 平均队列长度 | 吞吐量 |
| 队列1() | 56007.26 | 279976.8 | 666808.98 |
| 队列2() | 148482.52 | 742254.46 | 666808.98 |
| 队列3() | 194719.88 | 973391.9 | 666808.60 |

可以看出在饱和状态下，三队列均出现了队列长度无线增长的情况，三队列的吞吐量大小相等，在等待时间分布图中，三队列曲线的相对位置顺序也与队列值大小一致，证明我们的猜想正确。

1. 设置三队列的,，发包总数与轮询一次可处理包长与(2)相同，探究平均包长的变化对队列长度分布与等待时间分布的影响，三队列队列长度分布与等待时间分布如下所示：



,队列长度分布图



,等待时间分布图

模拟仿真得到的其他指标如下所示，服务台利用率（占用率）为76.50%：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 平均等待时间 | 平均队列长度 | 吞吐量 |
| 队列1() | 0.27 | 2.73 | 333580.07 |
| 队列2() | 0.29 | 2.93 | 277983.39 |
| 队列3() | 0.32 | 3.12 | 238271.46 |

可以看出，三队列的各项指标比较接近，只有在较小概率时三队列的等待时间分布与队列长度分布才会出现差异。同时注意到三队列在吞吐量上与值成反比，且由于三队列均处于不饱和状态，每个队列轮询的等待时间为另外两队列的到达包处理时间之和，由于包到达间隔相近，则每个队列在一次轮询的过程中处理的包的数量是相近的，在理想状态下进行分析，设一次轮询中每个处理的包的数量均为a，则每个队列在一次轮询中等待的时间如下：

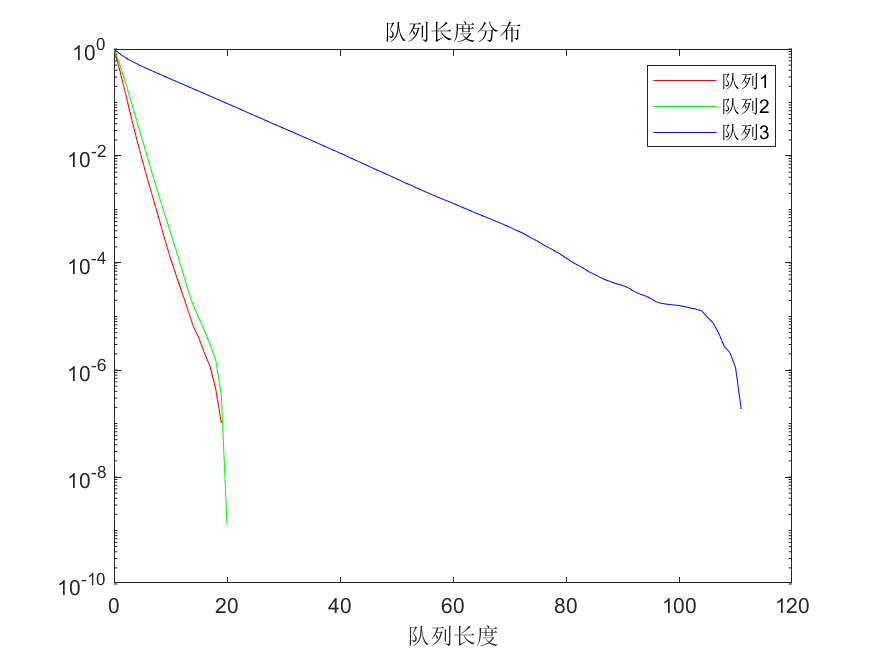
一次轮询中队列1等待时间

一次轮询中队列2等待时间

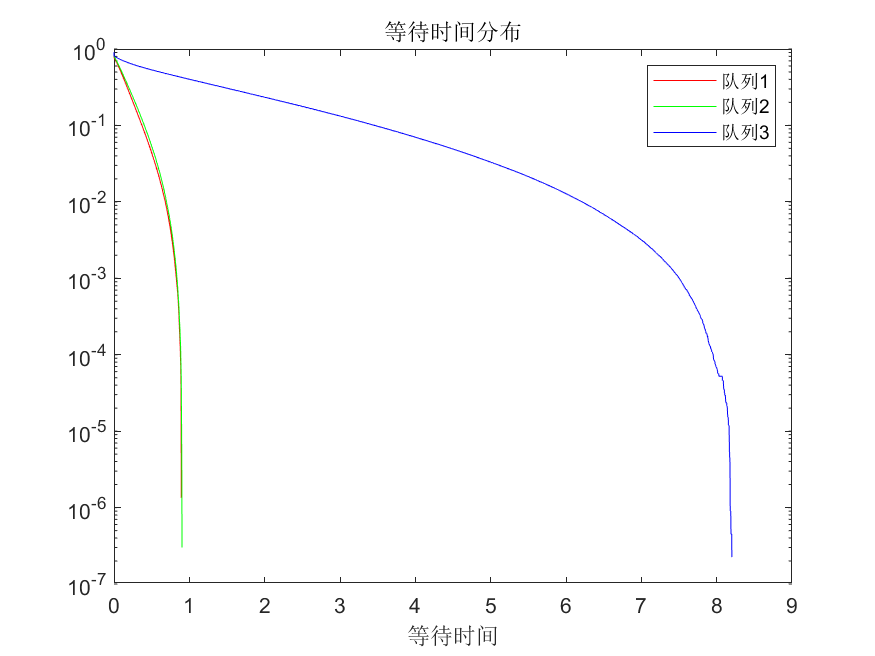
一次轮询中队列3等待时间

平均包长最长的队列1等待时间显然要小于其他队列的等待时间，因此在等待时间分布图中的曲线相对位置顺序即为平均包长大小的顺序，这也在一个侧面印证了(2)中的想法。用类似的原理可以解释为什么平均队列长度与平均等待时间的大小与队列平均包长大小相反。

(5)为了研究轮询一次可处理包长大小对等待时间分布与队列长度分布的影响，设置轮询一次可处理包长大小为0.1，其他参数与(2)相同，通过与(2)的对比，得出相应的结论，三队列队列长度分布与等待时间分布如下所示：



可处理包长为0.1，,队列长度分布图



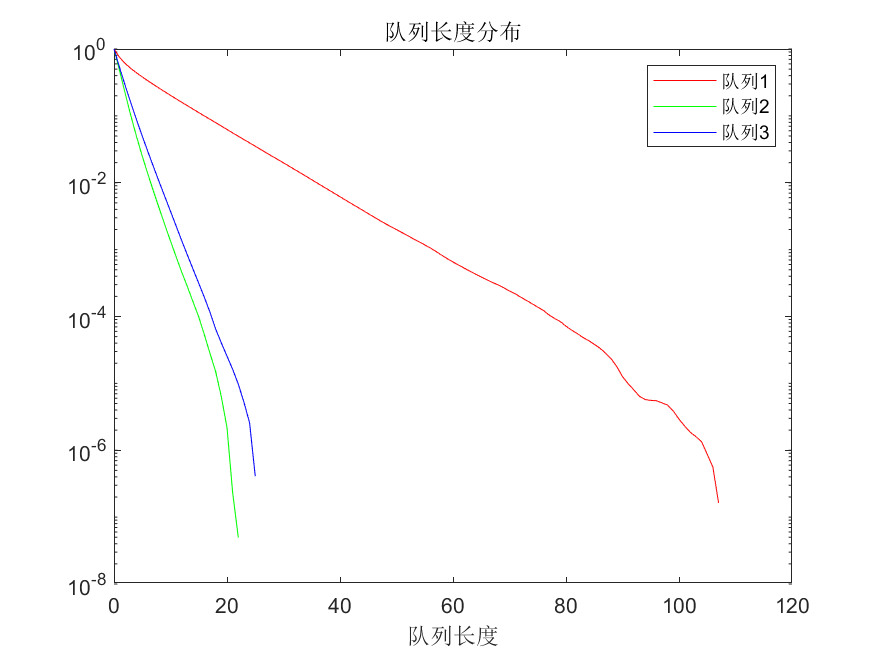
可处理包长为0.1，,等待时间分布图

模拟仿真得到的其他指标如下所示，服务器利用率（占用率）为90.05%：

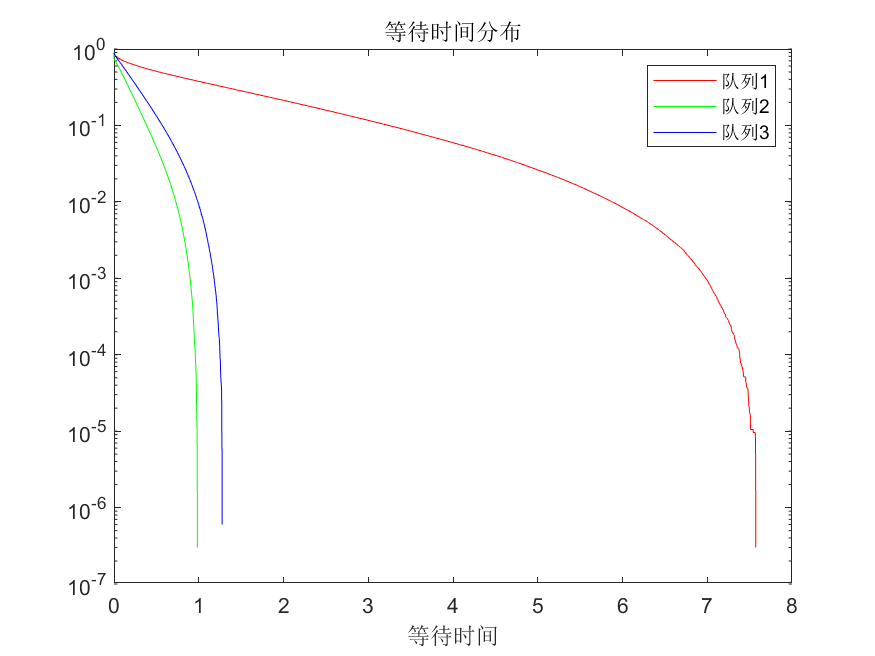
|  |  |  |  |
| --- | --- | --- | --- |
|  | 平均等待时间 | 平均队列长度 | 吞吐量 |
| 队列1() | 0.05 | 0.65 | 222320.21 |
| 队列2() | 0.07 | 0.95 | 333369.63 |
| 队列3() | 0.78 | 7.42 | 444690.26 |

可以看出，在上述情况下，分配较小的可处理包长对包到达频率低（即小）的队列更有利，平均等待时间与队平均列长度远小于包到达最频繁的队列。包到达更频繁的队列更易出现在一次轮询中无法处理完队列内所有包的情况，导致队列长度的积累，因此平均等待时间与平均队列长度都会更大。

为了进一步研究轮询一次可处理包长大小对等待时间分布与队列长度分布的影响，保持轮询一次可处理包长大小为0.1不变，其他参数与(3)相同，通过与(3)的对比，得出相应的结论，三队列队列长度分布与等待时间分布如下所示：



可处理包长为0.1，,队列长度分布图



可处理包长为0.1，,等待时间分布图

模拟仿真得到的其他指标如下所示，服务器利用率（占用率）为76.60%：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 平均等待时间 | 平均队列长度 | 吞吐量 |
| 队列1() | 0.59 | 5.63 | 333580.07 |
| 队列2() | 0.07 | 0.88 | 277983.39 |
| 队列3() | 0.10 | 1.16 | 238271.46 |

可以看出，在上述情况下，分配较小的可处理包长对平均包长小的队列更有利。因为平均包长更大的队列易出现在一次轮询中无法处理完队列中所有包的情况，导致队列长度的积累，因此平均等待时间与平均队列长度都会更大。