**《数据中心技术课程实验报告》**

|  |  |
| --- | --- |
| 姓 名： | 万贤林 |
| 专 业： | 电子信息 |
| 学 号： | M202173737 |
| 指导教师： | 施展、童薇 |

一、实验目的

1.使用对象存储服务器，观测在服务器上读写对象文件的吞吐率、延迟等性能指标。

2.观测长尾延迟，使用对冲请求缓解长尾延迟现象。

二、实验环境

|  |  |
| --- | --- |
| 服务器： | minio RELEASE.2021-11-24T23-19-33Z |
| 操作系统： | Windows 10 |
| 编程语言： | python 3.9 |

三、实验过程

**3.1 观察读写对象文件的性能指标（以写为例）**

实验中，准备一个大小为4096字节的文件，上传至对象服务器共100次，观察该过程中吞吐率和延迟。上传至服务器比特数为4096\*8=32768 bits，从开始上传至上传结束经历时间如图1所示，共3340.05ms.

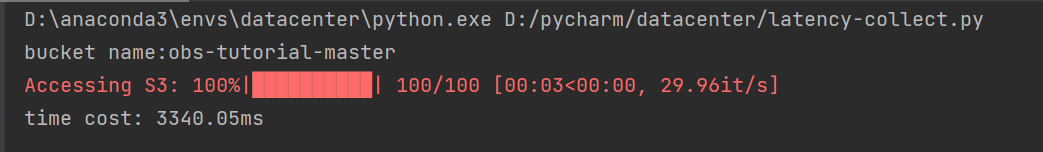


图1 上传所有文件总耗时

因此吞吐量计算公式为：

实验中，分别记录上传每个文件时的延迟latency，具体延迟信息如图2所示。

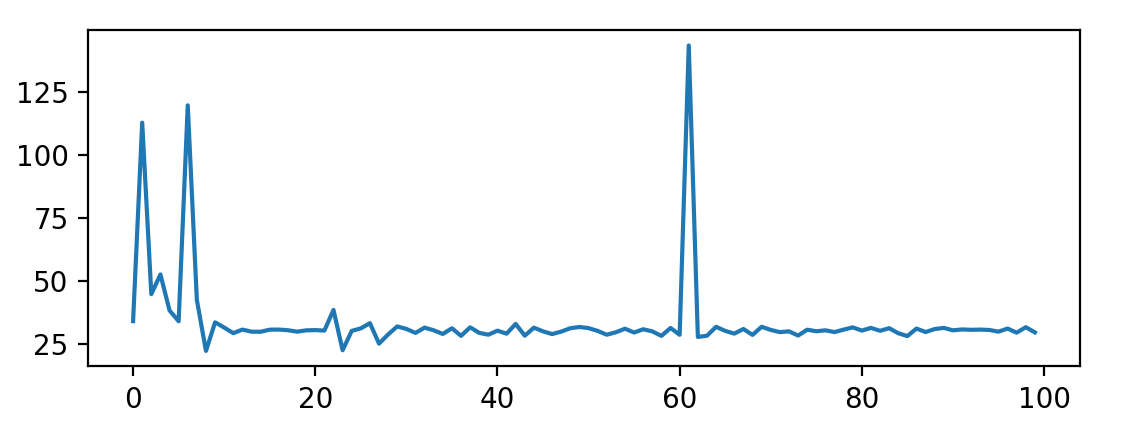


图2 各个文件上传延迟

将各个文件上传延迟降序排列绘制成图3.从图3中已经可以初见长尾延迟的现象了。

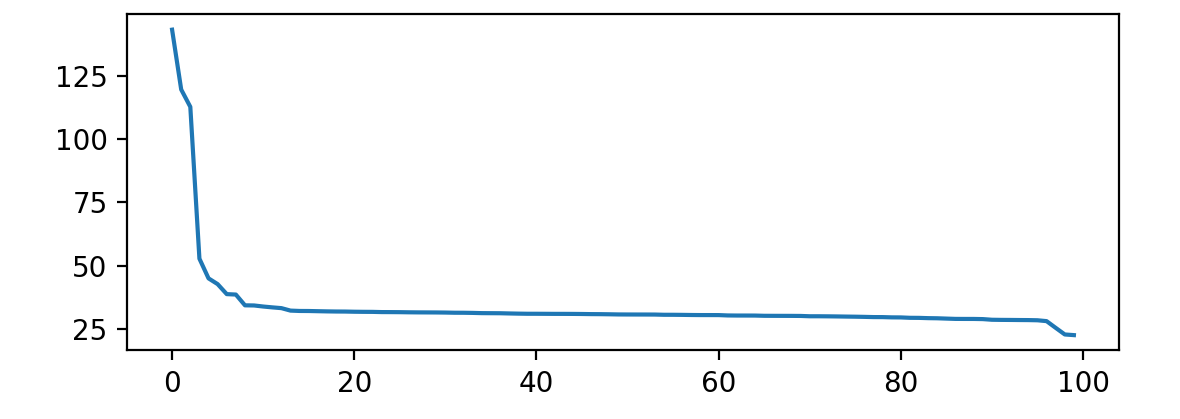


图3 各文件上传延迟降序排列

**3.2 观测长尾延迟现象**

为使长尾延迟更容易被观测，对收集到的延迟信息绘制直方图，如图4所示。可以明显观测到，超过90%的延迟都在（20ms，40ms）范围内，但仍有少量延迟远超40ms，最大延迟甚至超过140ms.

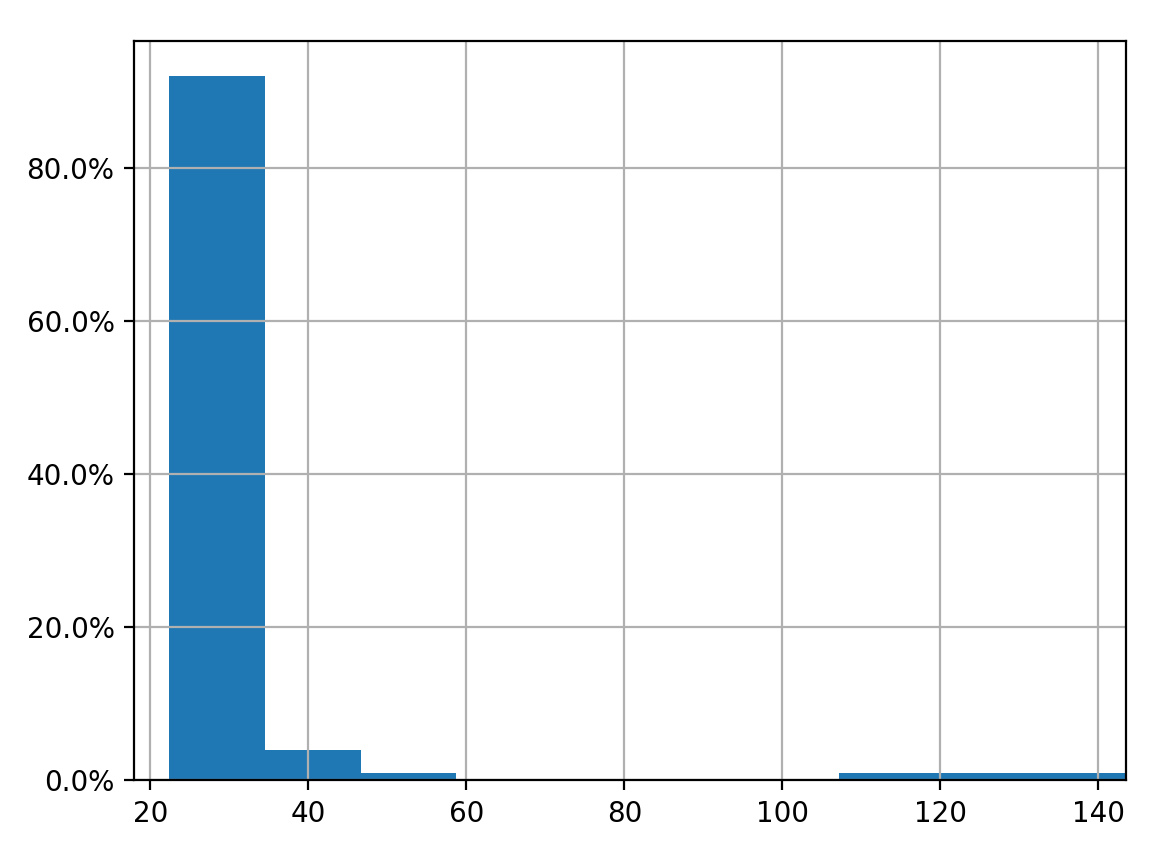


图4 延迟直方图

由于实验环境所限，只采集了100个延迟信息，为了模拟出真实应用场景下海量请求的延迟，使用排队论根据已有的延迟信息模拟绘制当有大量请求时的延迟曲线，如图5所示。

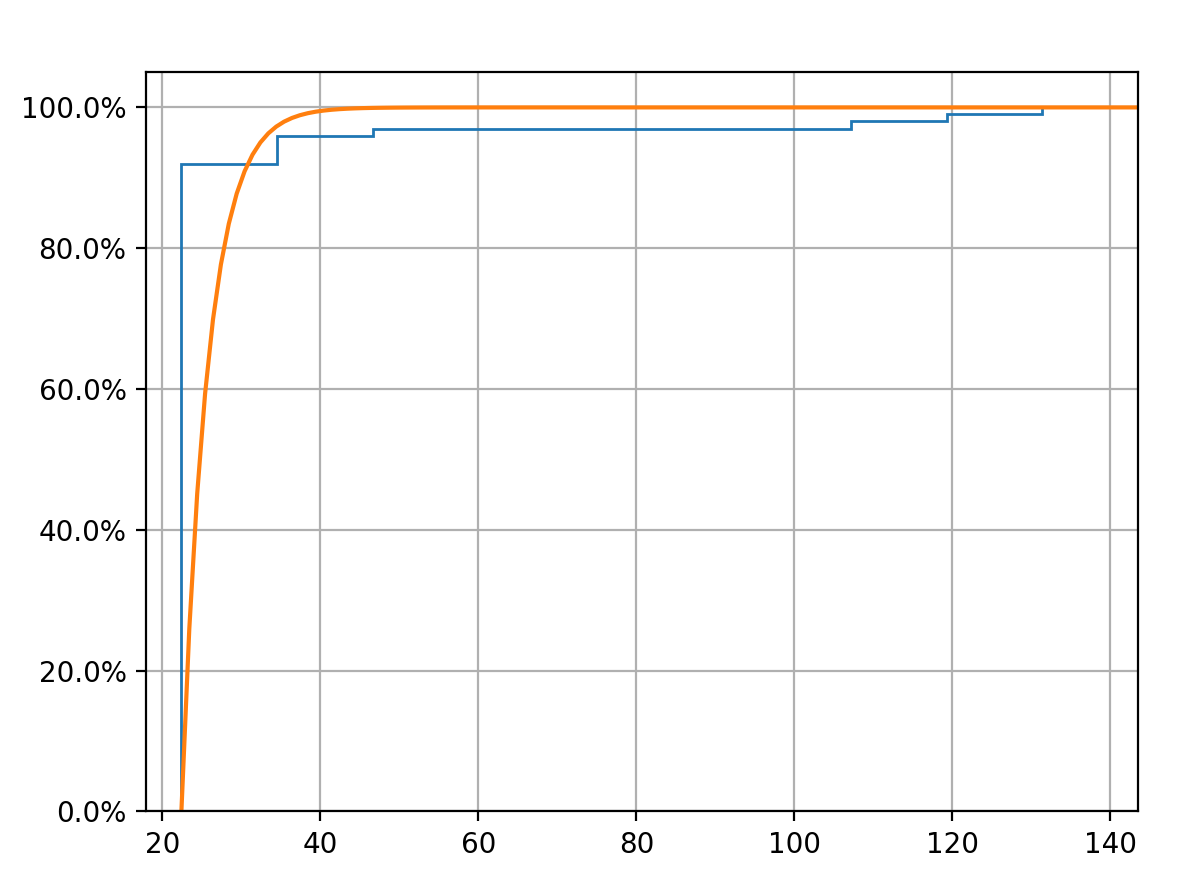


图5 使用排队论模拟延迟

**3.3 对冲请求**

为缓解长尾延迟，实验中采取的对冲策略为：当一个请求发出后30ms还未收到响应时，则重新发送该请求。采用该策略的原因是通过观测100个样本延迟，发现大部分请求都能在30ms内收到响应。若未在30ms内收到响应，则很有可能落入图4中（100ms，140ms）区间，所以应在30ms重新发送请求，这样就大概率能在60ms时收到响应，在一定程度上缓解了尾延迟现象。

采用对冲请求后绘制直方图如图6所示（与图4对比，减轻了尾延迟）。

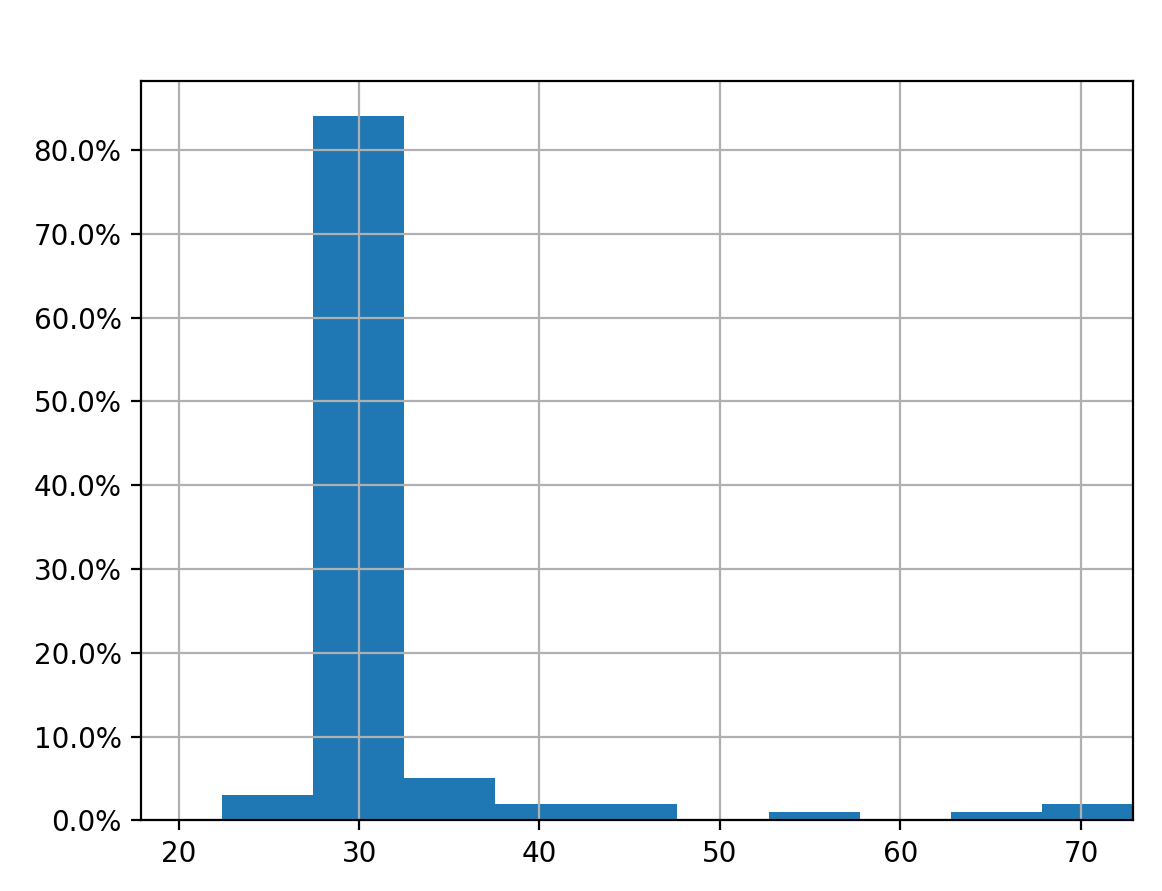


图6 使用对冲请求延迟分布直方图

采用对冲请求后，排队论模拟曲线如图7所示。

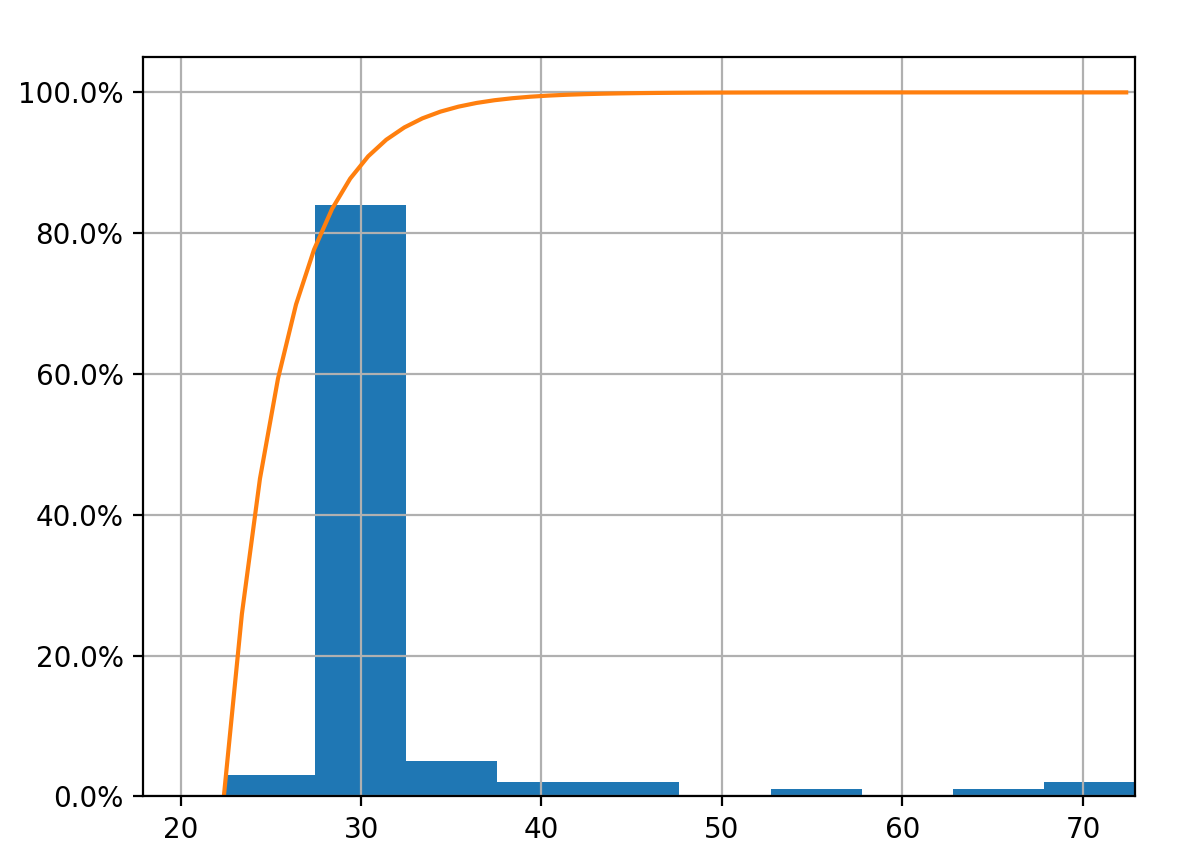


图7 使用对冲请求后延迟模拟曲线

四、实验总结

通过本次实验，熟悉了对象存储的基本概念，在实验过程中也了解了minio服务器的部分操作。通过亲身观测并使用对冲请求缓解尾延迟，对其有了更深的理解。