**计算机网络实验报告**

**基于 IP数据报分组的计算机网络层传输层研究**

姓名：张蕴琪、杨哲

学号：2015011229、2014010614

班级：无51，精41

1. **实验目的：**

通过利用 Tcpdump 工具熟悉计算机网络中网络层和传输层数据传输的过程和协议的使用。加深对于网络层中的 IP协议和传输层中的 TCP、UDP、ICMP等协议传输格式和与数据关系的理解。

1. **实验内容**

1. 用 Tcpdump 收集某个主机或者路由器所连接的某个物理网络上的traffic，存放到文件中以备进一步分析使用。收集流量的时间长短可以选择三种规格之一：A.5分钟；B.15 分钟；C.1 小时

2. 编写程序处理原始数据文件，整理成你认为方便处理的数据格式（纯文本）

3. 利用 Matlab 或其他工具，或自行编写程序，分别就进出两个方向上的 traffic，至少分析以下特征：

（1）给出 IP分组携带不同协议的载荷的饼图，分别按分组数和总数据量进行统计；

（2）有多少 IP 分组是片段（fragment）？有多少 IP 数据报被分片？载荷为 TCP何 UDP的分别有多少比例的 IP数据报被分片？

（3）给出 IP数据报长度的累积分布曲线，并分别比较载荷为 TCP何 UDP的 IP数据报长度的累积分布；

（4）分别对 TCP和 UDP的 traffic 给出端口分布的直方图，比较前10 名端口上数据报长度的累积分布曲线；

（5）对于载荷为 TCP 的报文，给出其中各个控制位出现的百分比。

4. 报告应包含原始数据用 Tcpdump 不带-x 参数读出的最初 1000行和最末 1000行作为附录。

1. **实验分工**

我们这一组一共有两名同学，张蕴琪同学负责前期数据的收集以及代码编写，杨哲同学负责后面实验报告的撰写。

1. **实验步骤与实验结果分析：**

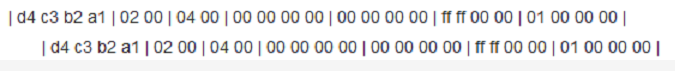
1、数据收集

在Windows系统下利用windump收集网络上的traffic，使用命令为：

windump -ip -i 1 -c 10000 -s 200 -w da.txt

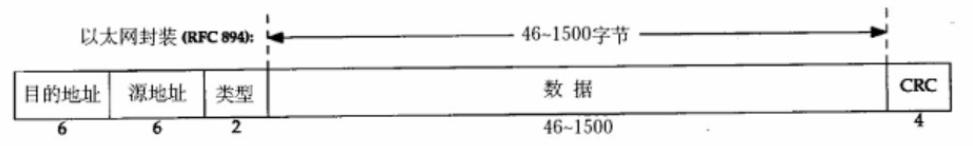
2、大作业实现思路

首先通过查询资料知道.pcap格式文件有如下文件头

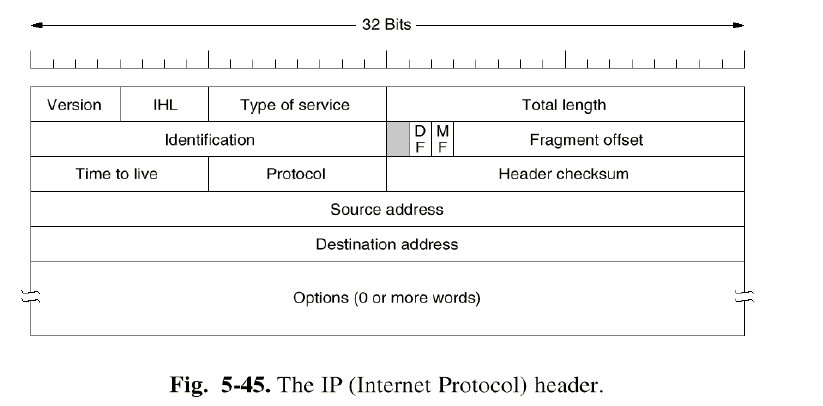


因此真正抓得的数据包从第25个字节开始， 之后的16个字节说明了抓取第一个包的时间，包的实际长度和抓取的包的长度，然后即分别为以太网帧包头、ip数据报包头、TCP或UDP等数据报包头。通过课上学习的知识可以知道以太网帧包头、ip数据报包头、TCP或UDP等数据报包头如下所示：

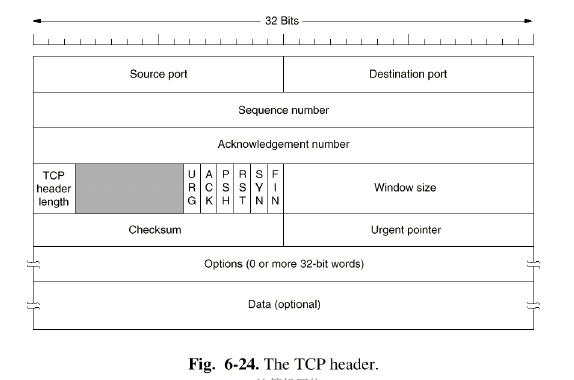
以太网帧包头：



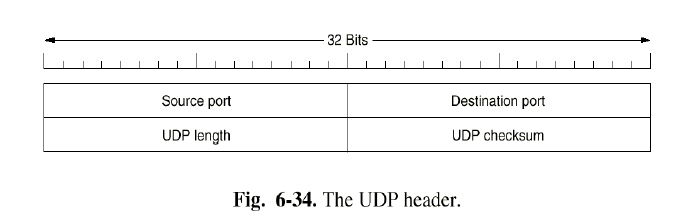
ip数据报包头：



TCP数据包头



UDP数据包头：



因此可以通过上述格式分别获取每一个抓取的数据包包头信息，从而完成这个大作业。

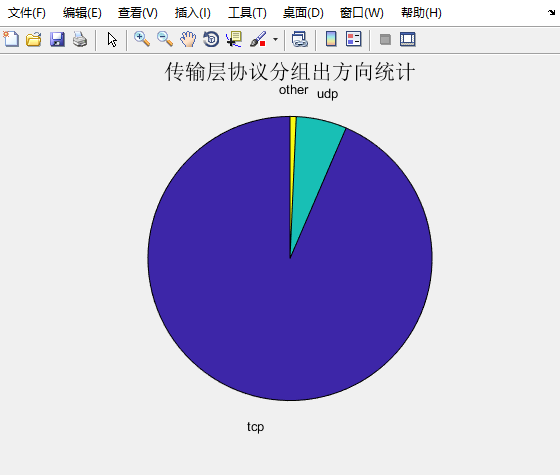
3、数据处理说明

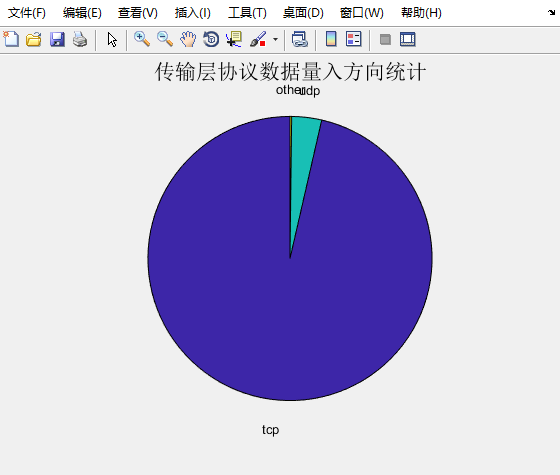
首先我们小组在采集完数据之后对得到的数据使用matlab语言进行处理，数据预处理主要代码在work.m文件之中，处理完毕之后得到了T1.mat,T2.mat,T3.mat,T4.mat,T5.mat数据文件。之后的代码(分别为pro1、pro3、pro4、pro5)利用T1-T5的数据进行数据统计和数据可视化工作，

4、

（1）直接利用数据预处理时候得到的Protocol位信息通过哈希表统计出入流量中TCP和UDP所占的百分比。（见下）其中出入流量的判定是通过对比本机MAC地址和源MAC地址得到的。TCP和UDP的辨识是通过Protocol信息位的值，06表示TCP，11表示UDP。

从图中可以看出，无论是分组数还是数据量，收发数据中TCP都占了绝大多数。代码实现在pro1.m中





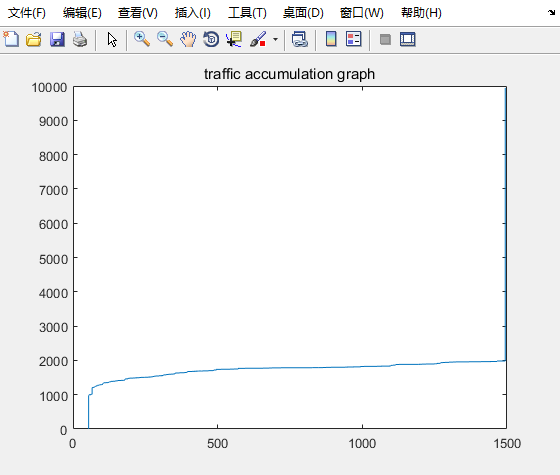
(2)

Ip包头中的DF和MF为表明了其是否分段，从T2.mat的数据来看，所有的包DF都设置为1或者是DF和MF都为0，说明数据报并没有被分组传输。可能是因为传输的数据量较小或者是Windump对监听到的分组进行了分割处理，所以没有被分组传输

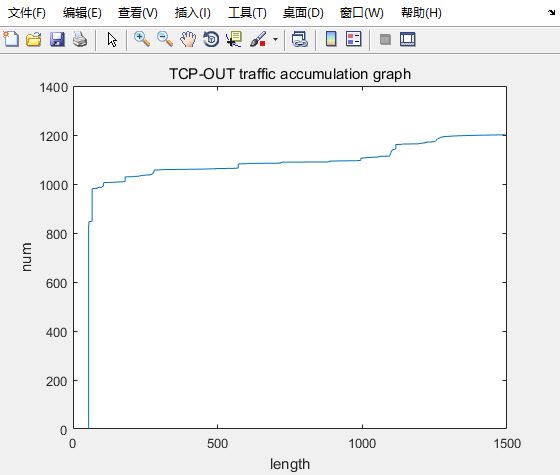
(3)

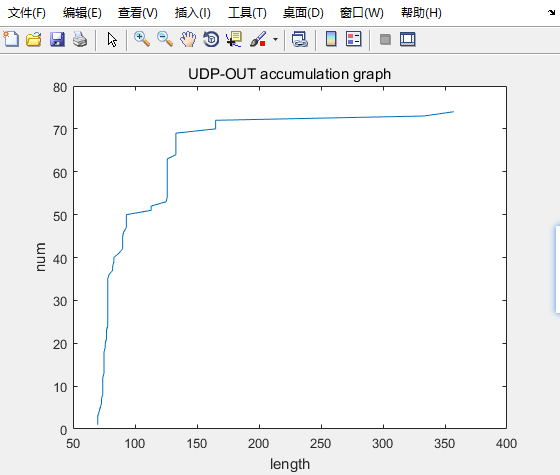
利用之前处理得到的T3.mat，运行pro3.m可以得到累进分布曲线。结果如下所示：

总的累进分布曲线：

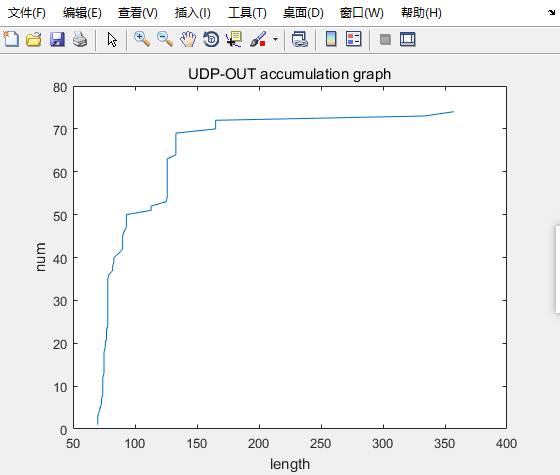


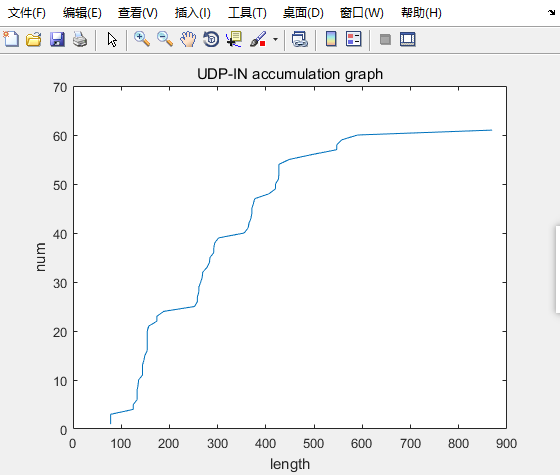
出方向累计分组统计：





入方向累计分组统计：



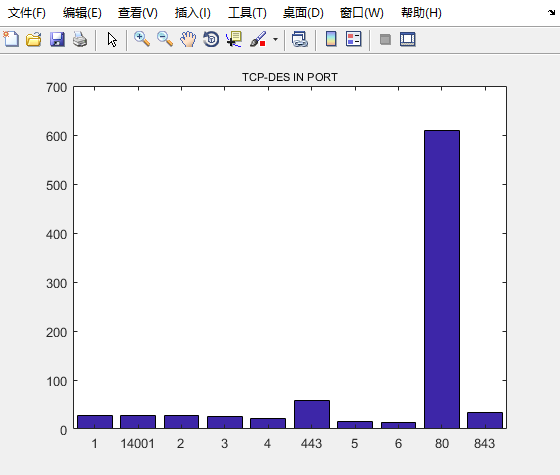
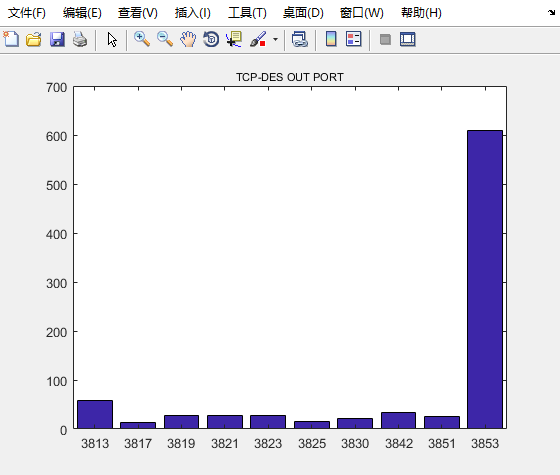


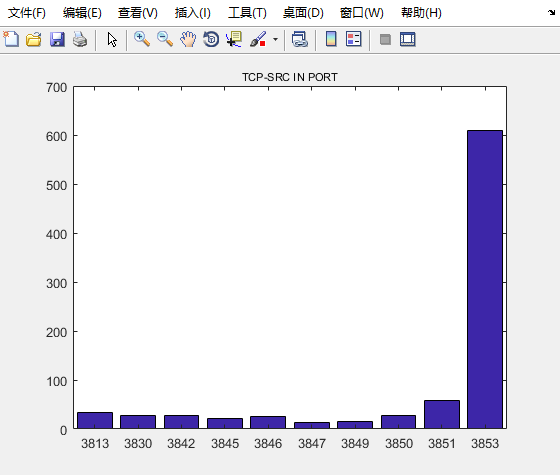
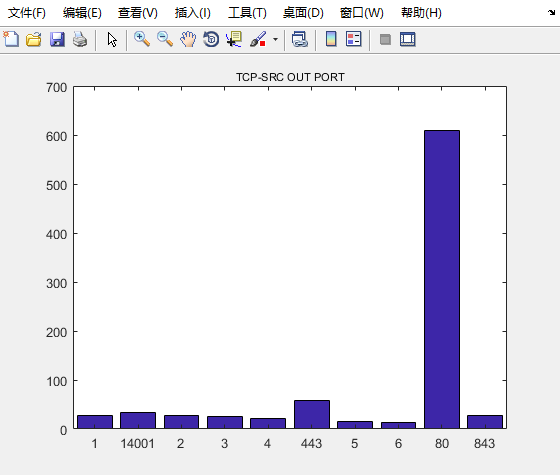
通过分析可以发现IP总数据报长度分布曲线和TCP数据报长度分布曲线更接近，这和（1）中分析结果一致，即在IP数据报分组传输的过程中绝大多数数据类型为TCP协议下的分组，因此 TCP 作为面向连接的分组传递方式占据主导地位。另外对比TCP和UDP的数据报长度分布曲线可以发现其中一段时间内TCP数据报长度几乎不变而UDP则呈线性增长，考虑到TCP对应网络上向用户进行文字、图像的信息传输。而 UDP 作为无连接的传输协议，在采样时间段内对应处理实时视频播放，因此出现了这种现象。

(4)

利用之前处理得到的T4.mat，运行pro4.m可以得到出入方向的最大端口统计：

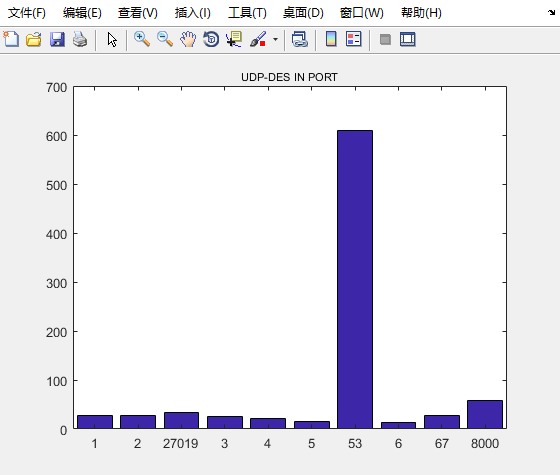
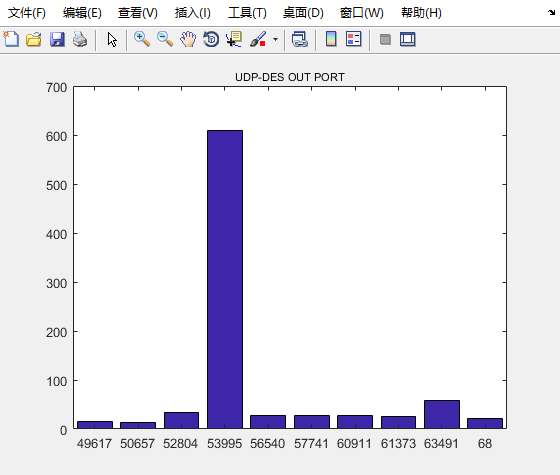
TCP端口：

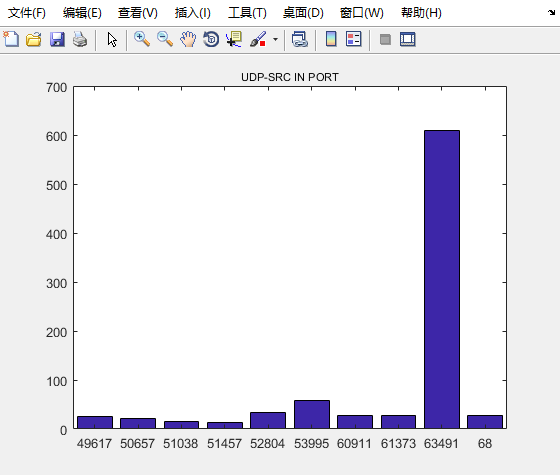
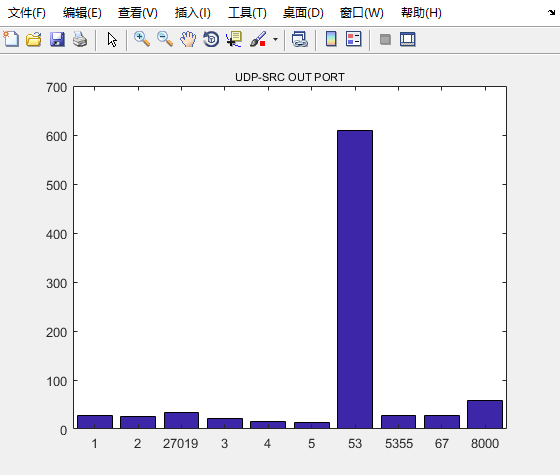




从上述端口中可以看出，本机进行TCP通讯的端口主要集中在80(HTTP服务)和443(HTTPS服务)的端口，这正好符合计算机网络的知识。

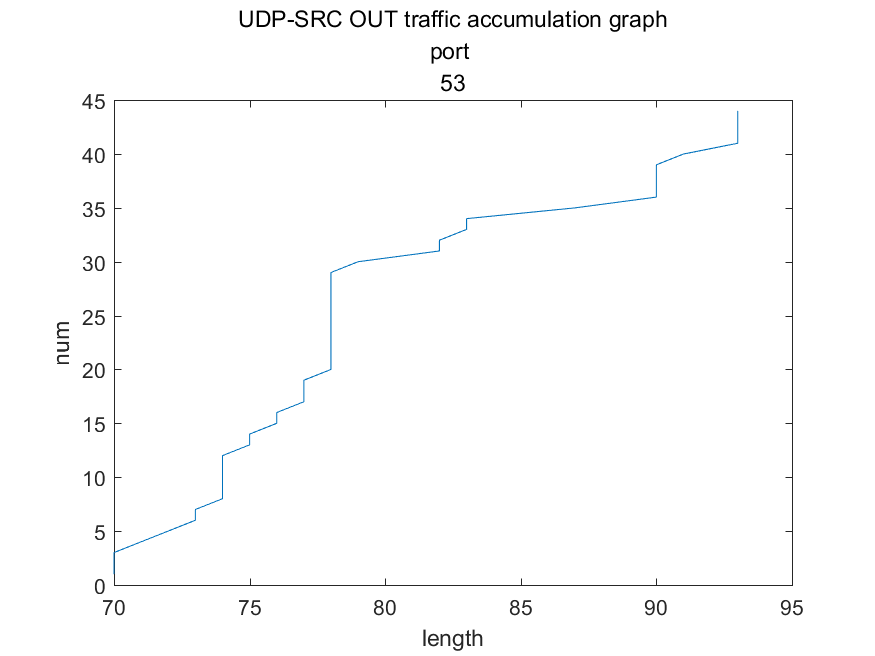
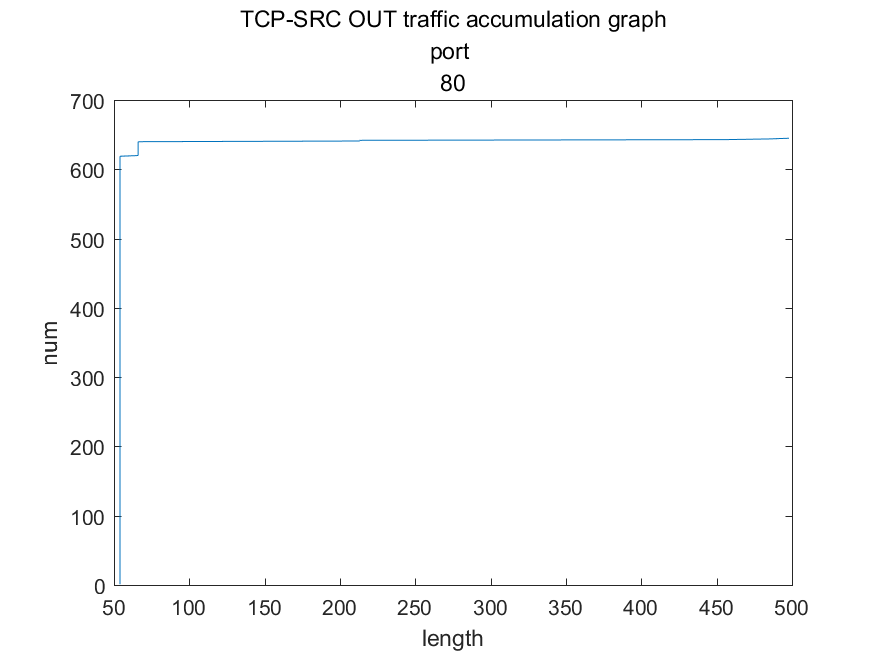
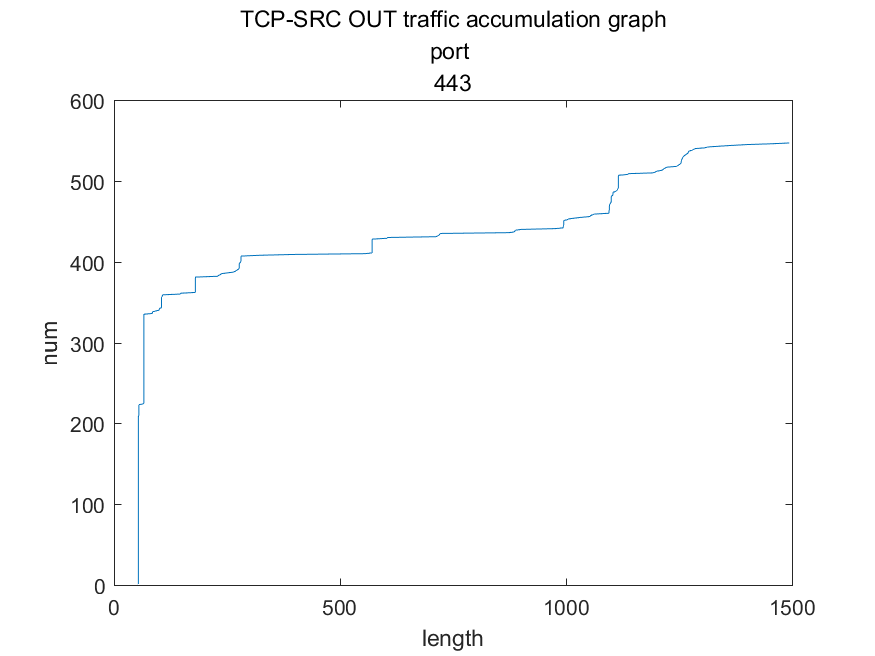
UDP端口：

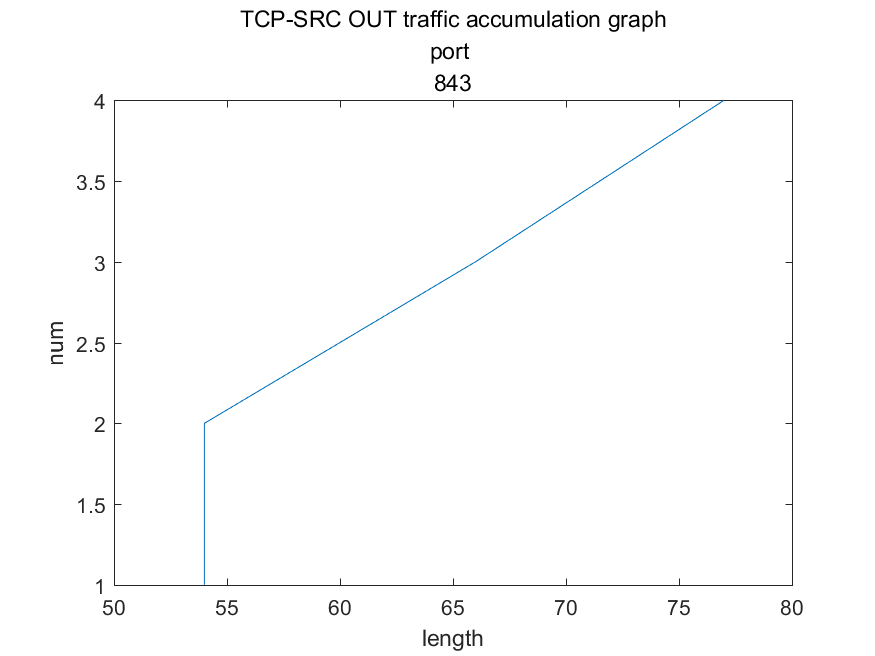




从上述图表中可以看出本机进行UDP通讯的端口主要集中在53(DNS服务)端口和67(DHCP服务)端口，这正好符合所学的计算机网络的知识。

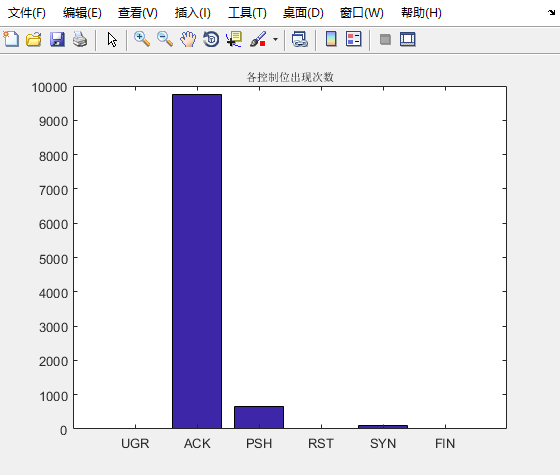
下面是一些端口的累计分布曲线(完整的在附件中所带的photo文件夹下)：

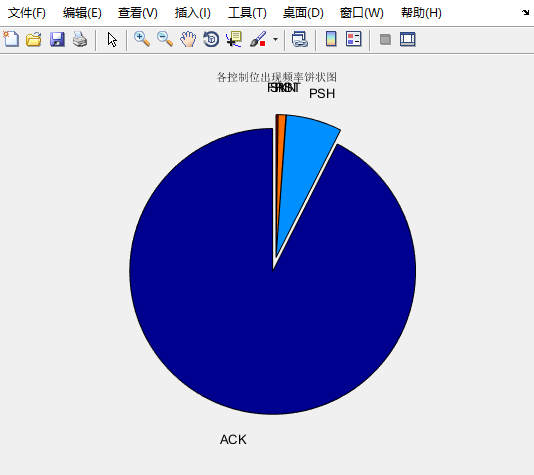




(5)

这一问实现的关键是准确确定每个控制位的位置和表示的信息。我们首先利用 Protocol 信息位确定筛选出属于 TCP 协议的数据报分组；然后监测每一个控制位上面的数据进行统计计数即可。利用之前处理得到的T5.mat，运行pro5.m可得如下结果为：





1. **实验感想：**

这次大作业实验我们收获很大。在课上，我们学习了网络层，传输层的功能以及网络层中的IP协议，传输层中的TCP和UDP等协议，然而在知道这些协议思想的同时对它们的具体实现方式却不是很了解。这次实验中，我们通过使用tcpdump命令收集数据并使用matlab分析这些数据，从更为实际的角度加深了对IP协议和IP头格式以及TCP、UDP协议的了解，也锻炼了我们获取网络数据的操作能力和编程处理能力。这次实验中，我们观察到TCP和UDP协议是使用最广泛的两种传输层协议，遍布了几乎所有的互联网活动，其中TCP由于可靠性高，被大量用于网页访问，文件传输等应用，也是使用数据量最多的传输层协议，而UDP协议不需要进行连接，传输延迟小，也具有一定的优势，在视频传输和语音传输等流媒体应用中应用广泛。

1. **实验原始数据：**

实验原始数据为da.txt，可在<https://github.com/zhang-2qi/computer_networks>中查询

数据的前1000行和最后1000行存放在附录.txt之中。