



北京邮电大学

Beijing University of Posts and Telecommunications

正则表达式应用

蒋砚军 北京邮电大学计算机学院



► 四个例题

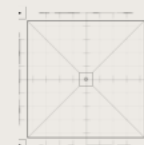


例1 累加作品发行量

例2 求一本书中出现频率最高的200个单词

例3 从网页中抽取数据绘制游客流量随时间变化的曲线

例4 根据捕获的网络流量观察TCP发送窗口变化



► 例4 根据捕获的网络流量观察TCP发送窗口变化



文件upload.cap是在Windows上使用软件WireShark捕获的文件上载过程中的网络流量。

执行命令 `tcpdump -ttnr upload.cap`

```
1228623354.154058 IP 192.168.0.108.2301 > 123.117.178.136.20: Flags  
[S.], seq 3210618050, ack 4098564053, win 65535, options [mss  
1460,nop,wscale 0,nop,nop,TS val 0 ecr 0,nop,nop,sackOK], length 0
```

```
1228623354.424039 IP 123.117.178.136.20 > 192.168.0.108.2301: Flags  
[.], ack 10081, win 6528, options [nop,nop,TS val 1620443362 ecr  
129980], length 0
```

```
1228623354.424059 IP 192.168.0.108.2301 > 123.117.178.136.20: Flags  
[.], seq 20161:21601, ack 1, win 65535, options [nop,nop,TS val  
129981 ecr 1620443362], length 1440
```

```
1228623354.424094 IP 192.168.0.108.2301 > 123.117.178.136.20: Flags  
[.], seq 21601:23041, ack 1, win 65535, options [nop,nop,TS val  
129981 ecr 1620443362], length 1440
```

期望得到的结果:

20.835455,5760
20.868127,5760
20.868165,7200
20.902815,7200
20.902853,8640
20.933327,8640
20.933365,10080
20.965792,10080
20.965829,11520
20.998753,11520
20.998788,12960
21.031269,12960
21.031318,14400
21.063662,14400
21.063710,15840



北京邮电大学

Beijing University of Posts and Telecommunications



谢谢