

一、实验目的

在这个项目中, 将使用两个命令行工具来分析互联网行为。你们将使用 `ping` 命令来测量 `RTT`(往返时间)和丢包率, `traceroute` 来检查网络路由。该实验的重点是研究100个你最常访问的网站(按 `Alexa` 衡量)的连接。在某些情况下, 我们会要求你对较小的一组网站进行更详细的测量, 以深入了解某些特定的测量。

二、提供的文件

`alexa_top_100`: `Alexa` 排名前 100 的最受欢迎的网站。

三、实验内容

1. `Ping` 命令

在这部分作业中, 主要使用 `ping` 来测量不同网站的往返时间。首先, 你的程序中应该具有一个函数: `run_ping`

1) `run_ping` 函数。应该运行 `ping` 命令并生成 `json` 输出。它应该有 4 个参数:

hostnames: 要 ping 的主机名列表;

num_packets: 发送给每个主机的 ping 数据包数;

raw_ping_output_filename: 将原始 ping 结果输出到(作为 json)的文件名;

aggregated_ping_output_filename: 将汇总的 ping 结果输出到(作为 json)的文件名。

例如,假设 `hostnames = [“www.baidu.com”]` 和 `num_packets = 100`,那么你应该 ping baidu.com 100 次。你可以通过使用子进程模块调用 `ping shell` 命令来执行此操作:

```
ping -n 100 baidu.com
```

注意:Windows 上的 -n 参数可能不同。如果不行,请尝试使用 -c。 `run_ping` 函数应该生成两个 json 输出。

1 原始ping结果。该文件包括每个数据包级别的详细结果。数据格式为:

```
{  
    hostname1: [rtt_1, rtt_2...],  
    hostname2: [rtt_1, rtt_2...],  
    ... }
```

每个主机名映射到 RTT 列表。按照上一个例子,如果你 ping baidu.com 100 次,那么应该有一个主机名 “baidu.com” 映射到 100 个 RTT 的列表。主机名应为字符串,RTT 列表应为浮点数,以毫秒为单位。 请注意,并不是所有的网站 都会响应 ping,有些

可能丢弃数据包。如果 ICMP 数据包超时,则请使用-1.0 作为 RTT。

2 统计汇总ping结果。对于每个网站,你应该汇总原始ping结果以产生 中间值 RTT,最大 RTT 以及丢包率。格式应该是:

```
{
    hostname1: {"drop_rate": drop_rate1, "max_rtt":
max_rtt1, "median_rtt": median_rtt1},
    hostname2: {"drop_rate": drop_rate2, "max_rtt":
max_rtt2, "median_rtt": median_rtt2},
    ...
}
```

主机名是字符串,并且每个映射到三个汇总数字:丢弃率,最大 RTT 和中值 RTT。应以 0.0---100.0 之间的百分比指定丢包率(例如:如果你观察到在 ping 运行中 500 个包中丢弃的数据包为 5,则丢包率为 1.0%)。最大 RTT 和中位数 RTT 应为毫秒。所有这三个数字应该是浮点数。如果一个网站有几个丢弃的 ping 数据包,请不要将这些数据包包含在中值/最大值计算中。如果一个网站根本没有响应,那么最大和中值 RTT 应该是-1.0,丢弃率应该是 100.0%。样例输出,请查看 sample_ping.txt 和

sample_ping.json。文本文件显示 ping google.com 的原始文本 (sample_ping.txt) 输出 10 次, json 文件(sample_ping.json) 显示 json 格式的原始 ping 结果。

sample_ping.txt

PING google.com

(216.58.194.206):

56 data bytes

64 bytes from 216.58.194.206: icmp_seq=0 ttl=54 time=2.561 ms

64 bytes from 216.58.194.206: icmp_seq=1 ttl=54 time=12.351 ms

64 bytes from 216.58.194.206: icmp_seq=2 ttl=54 time=2.963 ms

64 bytes from 216.58.194.206: icmp_seq=3 ttl=54 time=5.848 ms

64 bytes from 216.58.194.206: icmp_seq=4 ttl=54 time=2.590 ms

64 bytes from 216.58.194.206: icmp_seq=5 ttl=54 time=5.180 ms

64 bytes from 216.58.194.206: icmp_seq=6 ttl=54 time=3.610 ms

64 bytes from 216.58.194.206: icmp_seq=7 ttl=54 time=3.741 ms

64 bytes from 216.58.194.206: icmp_seq=8 ttl=54 time=2.752 ms

64 bytes from 216.58.194.206: icmp_seq=9 ttl=54 time=12.958 ms

--- google.com ping statistics ---

10 packets transmitted, 10 packets received, 0.0% packet loss

round-trip min/avg/max/stddev = 2.561/5.455/12.958/3.749 ms

sample_ping.json

```
{"google.com": [2.561, 12.351, 2.963, 5.848, 2.59, 5.18, 3.61, 3.741, 2.752, 12.958]}
```

2)将文本结果进行可视化,y 轴的值应为 RTT 注:Java 语言图表绘制,推荐使用 jfreechart 开源库

(<http://www.jfree.org/jfreechart/>)

*3)使用编写的代码,运行以下实验。

ping每个Alexatop100网站10次。你应该将代码的输出存储在两个文件中。接下来,我们要更详细地检查几个网站的 ping 行为。网站是:baidu.com, microsoft.com,sina.com.cn,taobao.com。

ping 每个网站 500 次。将代码的输出存储在两个文件中。

2. Traceroute 命令

虽然 ping 对于识别端到端行为很有用,但是 traceroute 是一个可以为你提供 有关 Internet 路由的更详细信息的工具。特别是,traceroute 允许你跟踪从机器 到远程机器的整个路由。

本部分要求:你应该编写一个程序,与 Part1 类似该程序应该运行 traceroute 列表的网站并生成 json 输出。该文件还应该能够直接解析 shell traceroute 命令输出。该程序应包含两个函数:

(1)`run_traceroute(hostnames,num_packets,output_filename):`

用于在 主机名列表上运行 `traceroute` 命令,输出 `traceroute` 的结果,它包括三个参数:`hostnames:traceroute` 的主机名列表;
`num_packets`:发送到每一跳的数据包数量; `output_filename`:在哪里保存 `traceroute` 命令的结果。

(2)`parse_traceroute(raw_traceroute_filename,output_filename):`

此函数 应该能够从 `run_traceroute` 函数中获取输出,并写出 `json` 数据。 `raw_traceroute_filename`:从shell命令存储`traceroute` 输出的文件的名称; `output_filename`:存储输出 `json` 数据。

你可以通过使用子进程模块调用 `traceroute shell` 命令来跟踪网站: Windows 系统: `tracert baidu.com` Linux 系统: `traceroute baidu.com`

你将查看以下网站的路由行为:`baidu.com,taobao.com,www.csu.edu.cn`,以及一些你感兴趣的网站。将输出结果保存为 `json` 格式。

二、操作环境

- 1.操作系统: Mac OS
- 2.编写语言: Java
- 3.编译软件: Eclipse

三、概要设计

1. Ping 操作:

(1) 首先通过 Java 调用终端指令 (Linux 指令), 并将原始的终端返回值保存至文件 Ping.txt 中。

(2) 从 Ping.txt 文件中读取 Ping 结果, 并依次分析, 使用正则表达式将返回结果中每次得到的 RTT 保存下来, 并提取出丢包率、最大 RTT、中值 RTT。

(3) 不断从 Website.txt 文件中提取网站, 对每个网站 Ping 十次, Ping 完并处理后的结果分别存储在 Ping_original.json 和 Ping_statistics.json 中。

2. JFreeChart 数据可视化

(1) 将 100 个网站 Ping10 次的结果和 4 个网站 Ping500 次的所有 RTT 结果分别用柱状图和折线图将数据可视化。

(2) 用柱状图表示各个网站丢包率、最大 RTT、中值 RTT。

[Ping 解释]

Ping 是一个用于测试另一台主机是否可达的小程序，Ping 的实现虽多种多样，但基本原理都是：发送一份 ICMP Echo Request (Echo Request) 报文给另一台主机，并等待返回 ICMP Echo Reply (Echo Reply)。

我们通常所说的 Ping 通是指发送 Echo Request 后并在一定的时间范围内收到了目的端对该请求的 Echo Reply。这里需要强调的是：尽管 Ping 依次发出多个 Echo Request，但是他们的内容不是完全一样的，而是通过序列号区别的。同样，Echo Reply 与 Echo Request 一一对应。因此，就算 Ping 收到一堆的 Echo Reply，但是如果内容和序列号对不上同样得不到认同。

我们通常所说的 Ping 不通是指发送 Echo Request 后并在一定的时间范围内没有收到目的端对该请求的 Echo Reply，也收不到其他任何的 ICMP 错误报文。

定位问题时通常需要结合选项 Ping，Ping 选项功能说明如下：

- a: 指定 Ping 命令的源地址；
- c: 设置发送 ICMP ECHO_REQUEST 报文的数目，默认数量为 5 个；
- d: 打开所使用 socket 上的 SO_DEBUG 选项；
- f: 设置报文发送的过程中不分片，默认允许分片；
- h: 指定发送 Echo Request 报文的 TTL 值，默认为 255；
- i: 指定 PING 命令的发送接口，若不指定则在 IP 层处理时根据 FIB 表查找出接口；
- m: 设置发送下一个 ICMP 报文的等待时间，默认为 500ms；
- n: 不对目的主机作域名解析；
- p: 设置 Echo Request 报文的填充字节 (pad)；
- q: 除统计数字外，不显示其他的详细信息；
- r: 记录路由，设置所使用 socket 上的 RECORD_ROUTE 选项和 Echo Request 包中的 RECORD_ROUTE 设置，并且显示返回包中的路由缓冲区；根据该选项可以记录 PING 报文的传送路径。但是要考虑到中间设备可能不支持该 socket 选项；
- s: 设置 ECHO_REQUEST 报文的数据段的长度，默认长度为 56bytes；
- t: 设置等待 ECHO_REQUEST 报文响应的超时时间，默认为 2000ms；
- tos 指定发送 Echo Request 报文的 TOS 值，默认为 0；
- v 显示接收到的非 Echo Reply 报文，缺省是不显示；
- vpn-instance 指定 VPN-Instance 的名称；

3. Traceroute 操作:

(1) 首先通过 Java 调用终端指令 (Linux 指令), 并将原始的终端返回值保存至文件 Traceroute_original.txt 中。

(2) 从 Traceroute_original.txt 中读取数据, 进行分析, 将每次路由 IP 地址保存下来, 如果出现 * 则表明被拒绝, 返回 No Answer。

(3) 将分析截取后的数据保存在 Traceroute.json 中。

[Traceroute 解释]

通过 traceroute 我们可以知道信息从你的计算机到互联网另一端的主机是走的什么路径。当然每次数据包由某一同样的出发点 (source) 到达某一同样的目的地 (destination) 走的路径可能会不一样, 但基本上来说大部分时候所走的路由是相同的。linux 系统中, 我们称之为 traceroute, 在 MS Windows 中为 tracert。traceroute 通过发送小的数据包到目的设备直到其返回, 来测量其需要多长时间。一条路径上的每个设备 traceroute 要测 3 次。输出结果中包括每次测试的时间 (ms) 和设备的名称 (如有的话) 及其 IP 地址。

在大多数情况下, 我们会在 linux 主机系统下, 直接执行命令行:

```
traceroute hostname
```

而在 Windows 系统下是执行 tracert 的命令:

```
tracert hostname
```

1) 命令格式:

```
traceroute[参数][主机]
```

2) 命令功能:

traceroute 指令让你追踪网络数据包的路由途径, 预设数据包大小是 40Bytes, 用户可另行设置。

具体参数格式: traceroute [-dflnrvx][-f<存活数值>][-g<网关>...][-i<网络界面>][-m<存活数值>][-p<通信端口>][-s<来源地址>][-t<服务类型>][-w<超时秒数>][主机名称或 IP 地址][数据包大小]

3) 命令参数:

- d 使用 Socket 层级的排错功能。
- f 设置第一个检测数据包的存活数值 TTL 的大小。
- F 设置勿离断位。
- g 设置来源路由网关, 最多可设置 8 个。

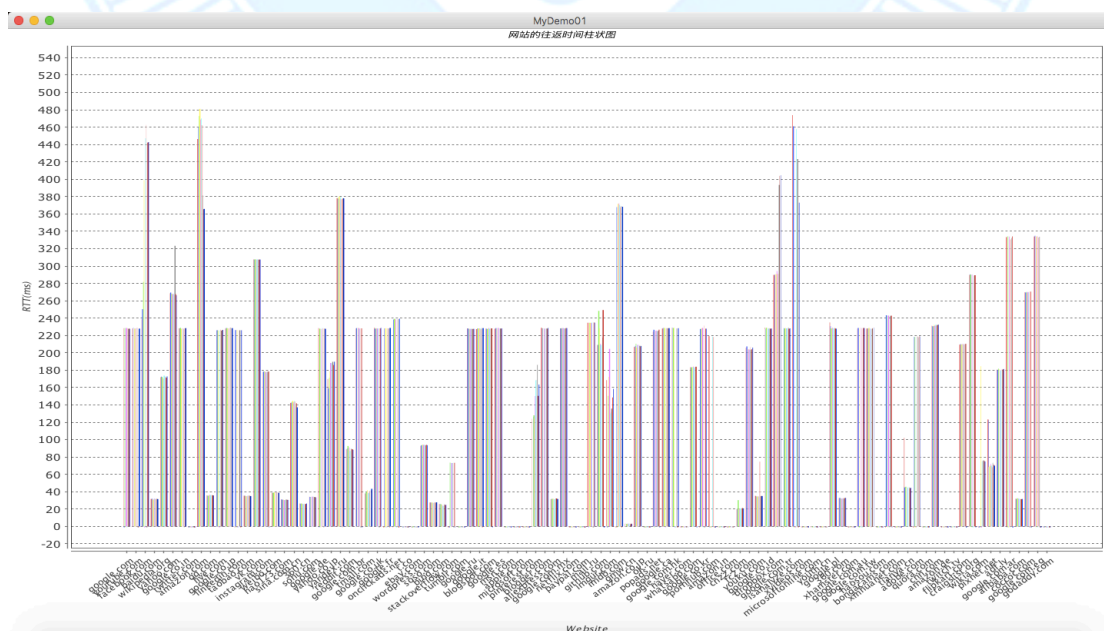
- i 使用指定的网络界面送出数据包。
- I 使用 ICMP 回应取代 UDP 资料信息。
- m 设置检测数据包的最大存活数值 TTL 的大小。
- n 直接使用 IP 地址而非主机名称。
- p 设置 UDP 传输协议的通信端口。
- r 忽略普通的 Routing Table，直接将数据包送到远端主机上。
- s 设置本地主机送出数据包的 IP 地址。
- t 设置检测数据包的特OS 数值。
- v 详细显示指令的执行过程。
- w 设置等待远端主机回报的时间。
- x 开启或关闭数据包的正确性检验

四、调试分析

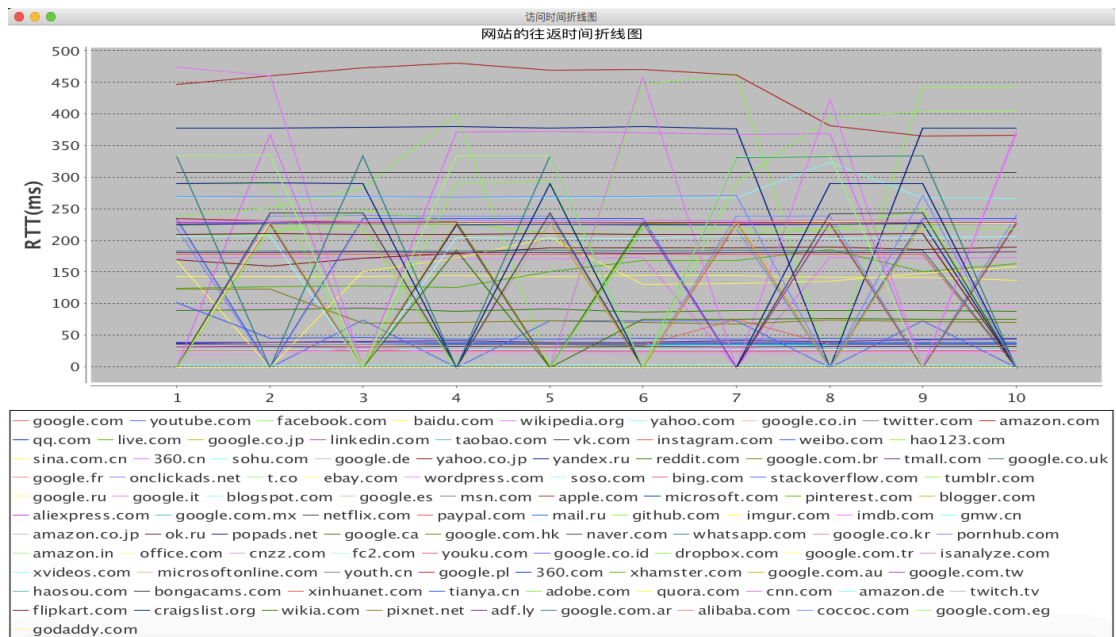
1. 100 个网站，每个 Ping10 次

(1) 原始结果、截取后的 RTT 结果、丢包率、最大 RTT、中值 RTT
见附件

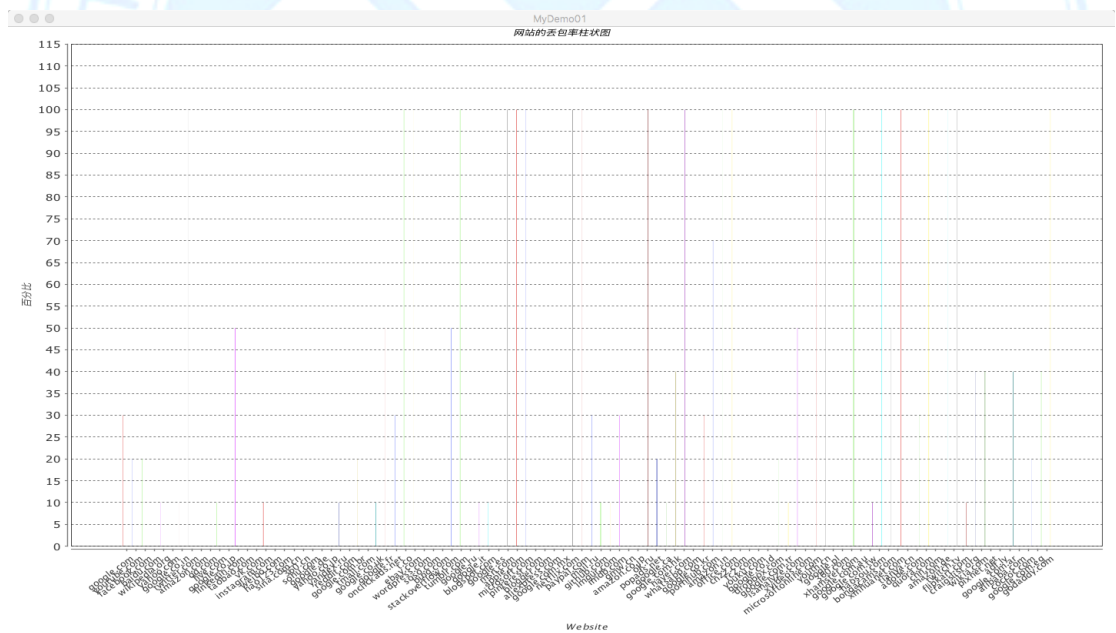
(2) RTT—柱状图



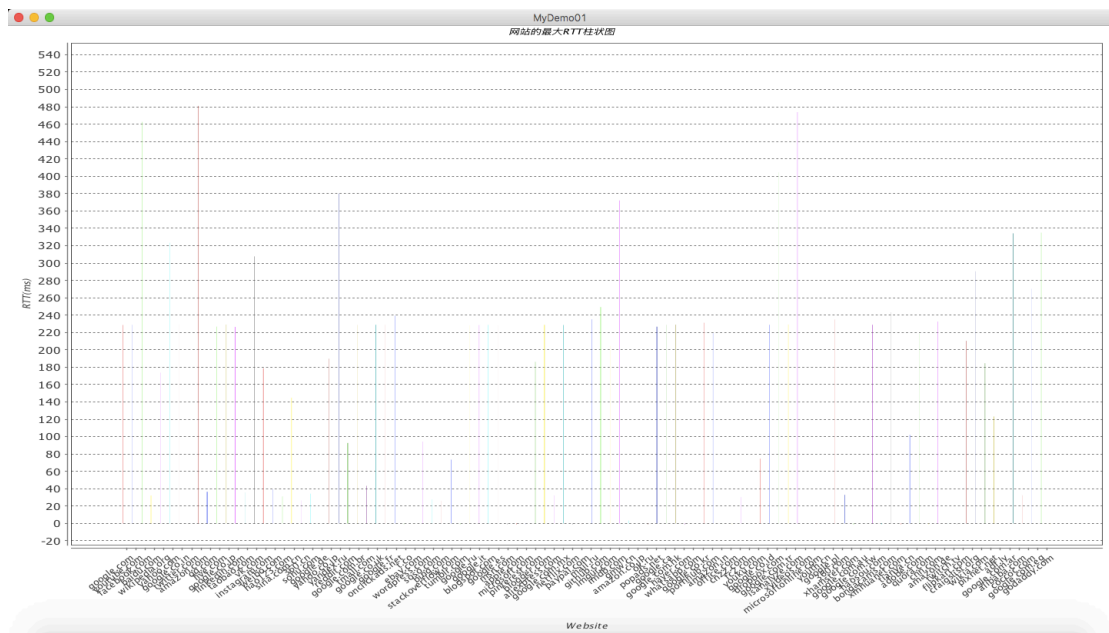
(3) RTT—折线图



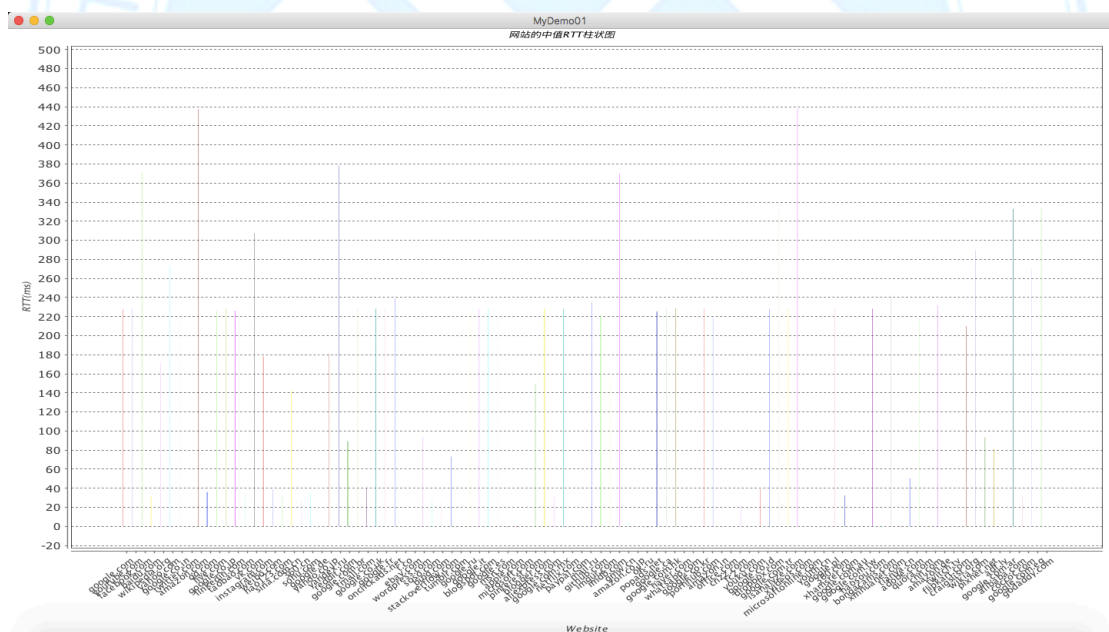
(4) 丢包率



(5) 最大 RTT



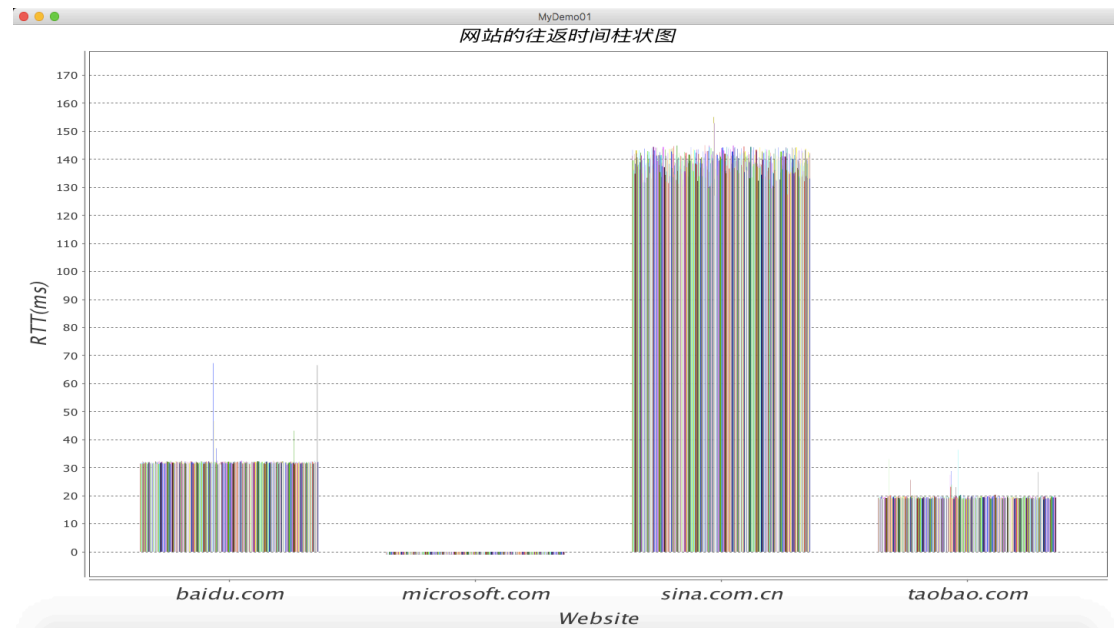
(6) 中值 RTT



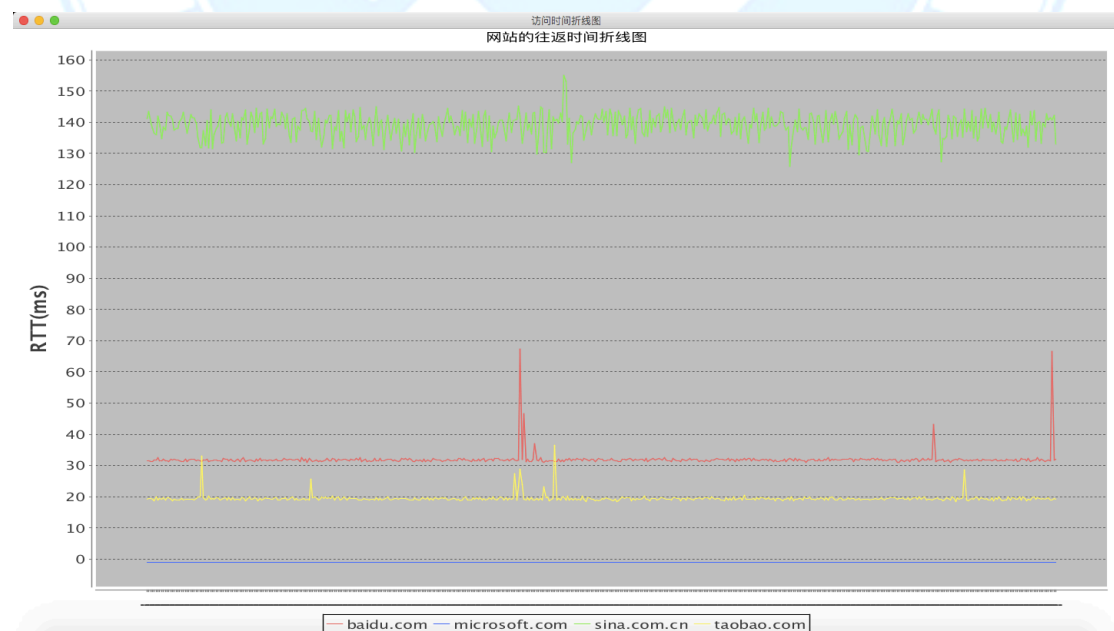
2.4 个网站，每个 Ping500 次

(1) 原始结果、截取后的 RTT 结果、丢包率、最大 RTT、中值 RTT
见附件

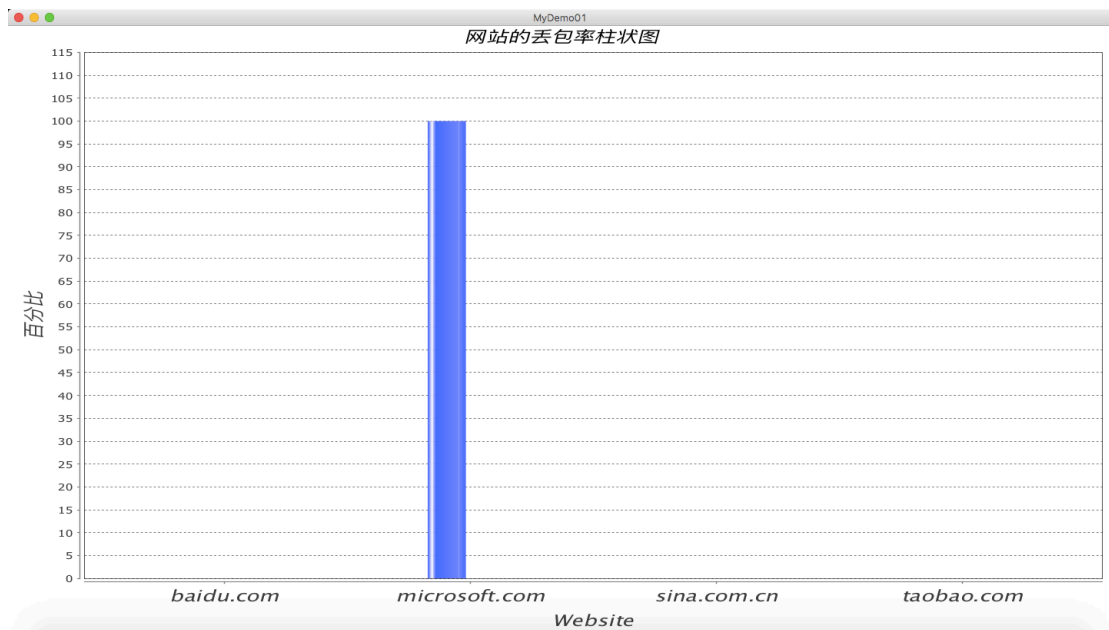
(2) RTT—柱状图



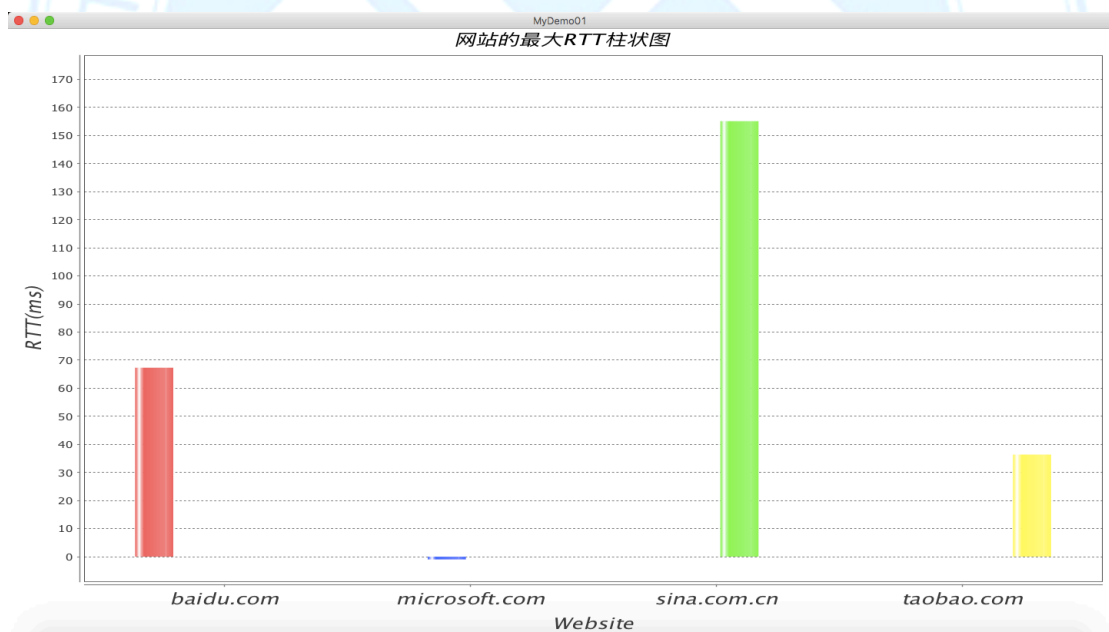
(3) RTT—折线图



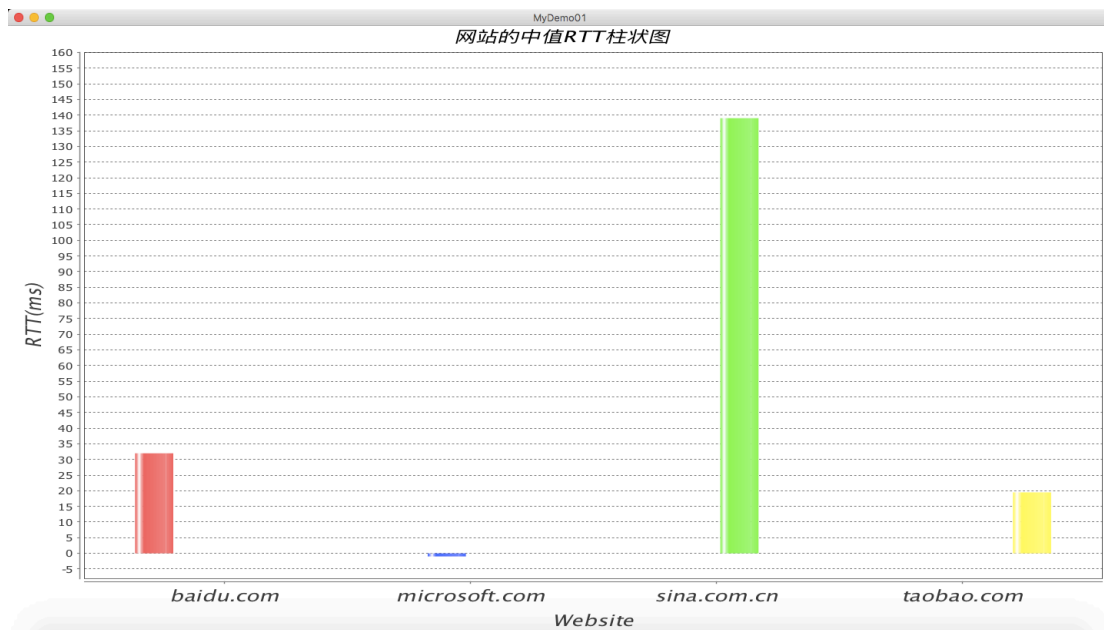
(4) 丢包率



(5) 最大 RTT



(6) 中值 RTT



3.Trace 指定的 3 个网站，每次 10 包

```
{
  "baidu.com":["10.96.0.1","222.247.30.157","61.137.3.249","No
Answer","220.181.177.70","No Answer","220.181.182.26","No
Answer","220.181.57.217"],

  "taobao.com":["10.96.0.1","222.247.26.101","61.137.9.221","No
Answer","220.191.200.106","115.236.101.217","42.120.247.93","
42.120.247.201","No Answer","No Answer","No Answer","No
Answer","No Answer","No Answer","No Answer","No
Answer","No Answer","No Answer","No Answer","No
Answer","140.205.94.189"],

  "www.csu.edu.cn":["10.96.0.1","222.247.26.97","222.247.29.130"
,"222.247.53.90"]
}
```

4.Trace 指定的 3 个网站，每次 3 包

```

{
    "baidu.com":["10.96.0.1","222.247.26.97","61.137.0.1","No
Answer","202.97.88.230","No      Answer","221.183.25.65","No
Answer","No      Answer","111.13.188.85","111.13.112.57","No
Answer","No Answer","111.13.101.208"],

    "taobao.com":["10.96.0.1","222.247.26.97","61.137.6.205","202.9
7.82.205","202.97.92.9","101.95.120.149","124.74.166.134","101.
95.208.14","101.95.211.126","No Answer","42.120.243.113","No
Answer","No      Answer","No      Answer","No      Answer","No
Answer","No      Answer","No      Answer","No      Answer","No
Answer","No      Answer","No      Answer","No      Answer","No
Answer","No      Answer","No      Answer","No      Answer"],

    "www.csu.edu.cn":["10.96.0.1","222.247.26.97","222.247.29.130"
,"222.247.53.90"]
}

```

4.Trace 自己喜欢的 3 个网站，每次 10 包

```
{ "google.com":["10.96.0.1","222.247.30.157","61.137.3.5","202.9
7.77.153","202.97.35.202","202.97.91.134","202.97.60.214","129.
250.2.155","129.250.7.39","129.250.7.21","116.51.18.194","113.2
1.245.94","113.21.245.194","113.21.241.233","122.144.25.118","
171.102.252.7","171.102.254.133","61.91.161.217"],

"github.com":["10.96.0.1","222.247.30.157","61.137.0.1","202.97.
77.57","202.97.60.1","202.97.91.146","202.97.58.226","202.97.90
.138","206.111.14.101","207.88.13.118","207.88.12.156","207.88.
12.158","207.88.13.97","64.3.225.14","No Answer","No
Answer","192.30.255.113"],

"youtube.com":["10.96.0.1","222.247.30.157","61.137.3.5","202.9
7.77.153","No
Answer","202.97.91.134","202.97.60.214","129.250.2.155","129.2
50.7.39","129.250.7.21","116.51.18.194","113.21.245.94","113.21
.245.194","113.21.241.233","122.144.25.118","171.102.252.7","1
71.102.254.133","61.91.161.217"]
}
```