

1 用户信息提取

根据“iviserver.log”日志文件中，每三条代表一个用户，我们可以通过源地址判断出每个用户为纯 v4、纯 v6 还是双栈用户，并且可以得到该用户的 IPv4 地址、IPv6 地址以及端口号，然后再在“client.log”日志文件中查询对应的操作系统便可以得到用户信息。在此列举其中几条用户信息如表1所列。

Table 1: 用户信息举例

用户类别	IPv4 地址	IPv6 地址	端口号	操作系统
纯 v4	106.11.156.54	None	58778	None
纯 v6	None	2001:da8:e265:0:c0a8:103::	55225	None
双栈	121.194.167.25	2001:250:3::73:4	55832	Win32

2 用户信息分析

2.1 纯 v4 用户、纯 v6 用户和双栈用户的分布比例

根据“iviserver.log”日志文件中，每三条代表一个用户，我们可以通过源地址判断出每个用户为纯 v4、纯 v6 还是双栈用户，然后统计纯 v4、纯 v6 还是双栈用户的数量如表2所示，我们绘制出三种用户分布的饼状图如图1所示，从图中可以直观的看出，大部分访问的用户均为纯 IPv4 用户，有少部分的用户既支持 IPv4 又支持 IPv6，只有非常少的用户是纯 IPv6 用户，占比不到 1%

Table 2: 纯 v4 用户、纯 v6 用户和双栈用户的数量及占比表

用户类别	数量	比例
纯 v4	1787385	0.92987
纯 v6	2418	0.00126
双栈	132388	0.06887

2.2 操作系统分布比例

根据“client.log”日志文件，统计所有的操作系统信息，绘制出操作系统使用最多的 5 种的饼状图如图2所示。从图中可以看出，使用 windows 操作系统的人数最多，超过了总记录的一半，这也与我们日常生活中使用 windows 系统的人数多有很大关系；linux 类的系统紧随其后，这可能是由于使用 android 系统智能机的人数较多；iphone 系统也进入了使用最多的操作系统的前五，这也可能是由于使用 iPhone 手机的用户较多。

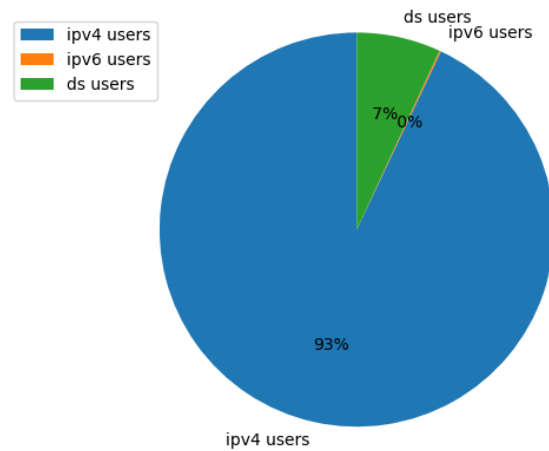


Figure 1: 纯 v4 用户、纯 v6 用户和双栈用户分布比例饼状图

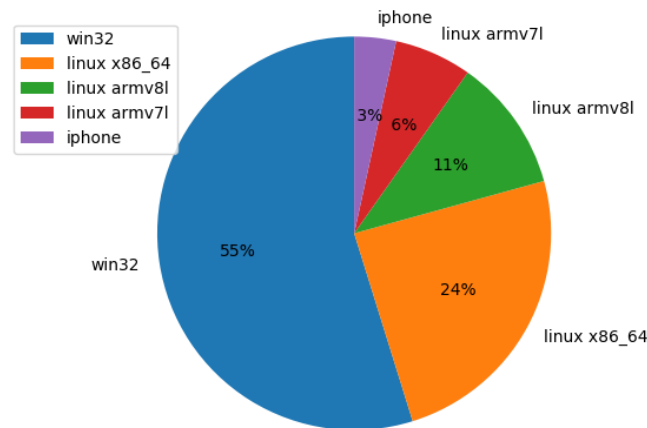


Figure 2: 使用最多的 5 种操作系统信息饼状图

2.3 各操作系统下纯 v4、纯 v6 和双栈用户分布比例

由于操作系统较多，因此我们在此选择使用人数最多的两种操作系统进行分析。首先我们对 win32 系统进行分析，纯 v4、纯 v6 还是双栈用户的数量如表3所示，绘制出用户分布饼状图如图3，从中我们可以看出，使用 win32 系统的用户中，依然大部分访问的用户均为纯 IPv4 用户，有少部分的用户既支持 IPv4 又支持 IPv6，只有非常少的用户是纯 IPv6 用户，占比不到 1%。

Table 3: win32 纯 v4 用户、纯 v6 用户和双栈用户的数量及占比表

用户类别	数量	比例
纯 v4	465450	0.75723
纯 v6	1007	0.00164
双栈	148221	0.24113

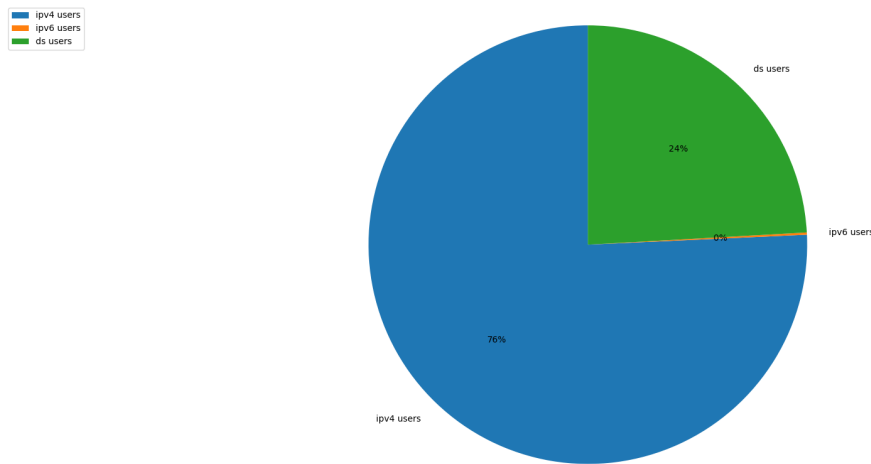


Figure 3: win32 系统纯 v4 用户、纯 v6 用户和双栈用户分布比例饼状图

其次我们对 iphone 系统进行分析，纯 v4、纯 v6 还是双栈用户的数量如表3所示，绘制出用户分布饼状图如图3，从中我们可以看出，使用 win32 系统的用户中，依然大部分访问的用户均为纯 IPv4 用户，有少部分的用户既支持 IPv4 又支持 IPv6，只有非常少的用户是纯 IPv6 用户，占比不到 1%。

Table 4: iphone 系统纯 v4 用户、纯 v6 用户和双栈用户的数量及占比表

用户类别	数量	比例
纯 v4	23280	0.64461
纯 v6	103	0.00285
双栈	12732	0.35254

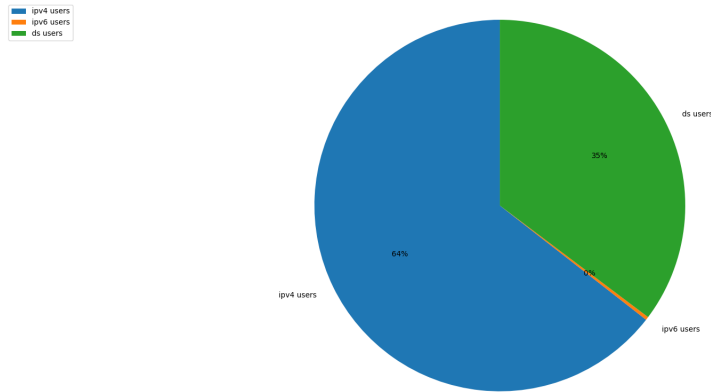


Figure 4: iphone 系统纯 v4 用户、纯 v6 用户和双栈用户分布比例饼状图

2.4 操作系统和端口号是否存在关联

此小节我们首先利用“client.log”日志文件得到操作系统以及端口号组成的二元组，然后我们将其绘制成散点图如图5所示，从图中我们可以看出，有些操作系统的端口号是连续的，例如：“linux aarch64”，“win64”等，而有些操作系统的端口号仅仅是几个离散的点，例如：“symbian”，“jva32”等。所以可以说明操作系统和端口号有一定的关联性。

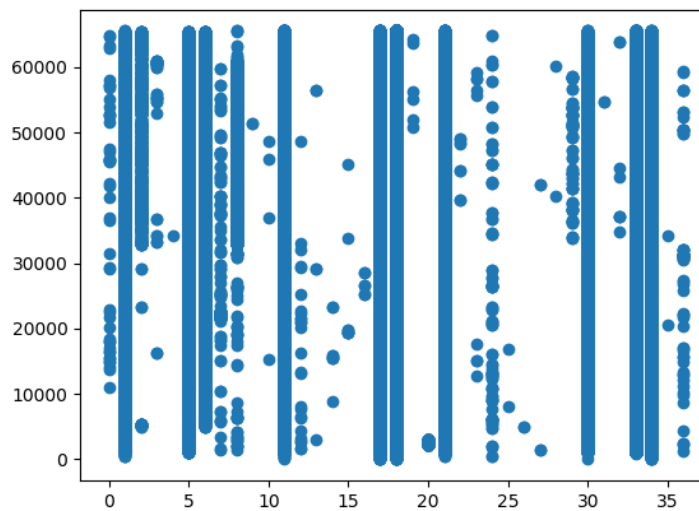


Figure 5: 操作系统和对应端口号的散点图

3 BGP 信息分析

首先我们使用爬虫工具https://github.com/gairehari/asn_info_crawler爬取了 AS 号信息，并将 AS 号信息存储在“as_info.json”文件中。然后根据 BGP 路由表得到的 ip 地址所对应的 AS 号查询运营商信息以及所在地域信息。首先我们绘制出 IPv4 地址中数量最多的前 10 大地域信息如图6所示，可以看出最多的地域分布在中国，这也与我们的正常认知相同。其次我们绘制出 IPv4 地址运营商中数量最多的前 10 大运营商的饼状图如图7，可以看出最多的运营商是“China Telecom Backbone”，紧随其后的是“China Unicom Backbone”并且这两者占据了绝大部分的运营商信息。

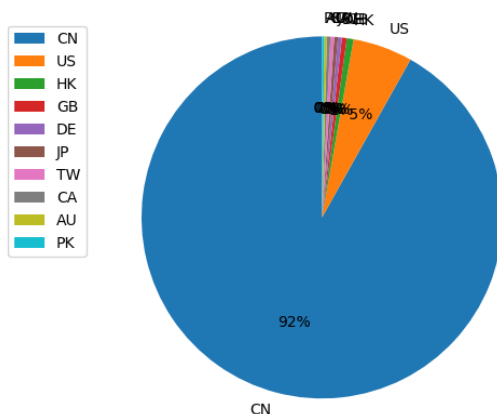


Figure 6: 最多的 10 个地域信息饼状图

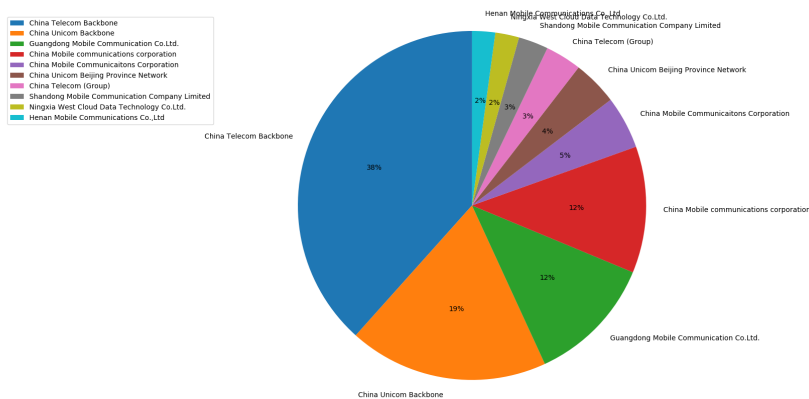


Figure 7: 最多的 10 个运营商信息饼状图

按照同样的方法，我们绘制出 IPv6 地址中数量最多的前 10 大地域信息如图8所示，可以看出最多的地域分布在中国，并且中国分布的比例比 IPv4 地址中国的比例还要高。其次我们绘制出 IPv6 地址运营商中数量最多的前 10 大运营商的饼状图如图9，可以看出最多的依然运营商是“China Telecom Backbone”，但是排第二的运营商变为了“Guangdong Mobile Communication Co.Ltd.”，并且排第二的运营商和排第三的运营商所占的比例非常接近。

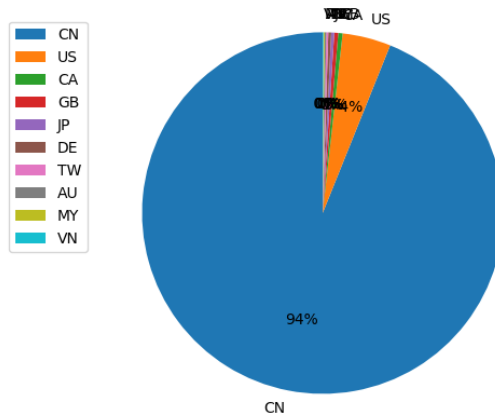


Figure 8: 最多的 10 个地域信息饼状图

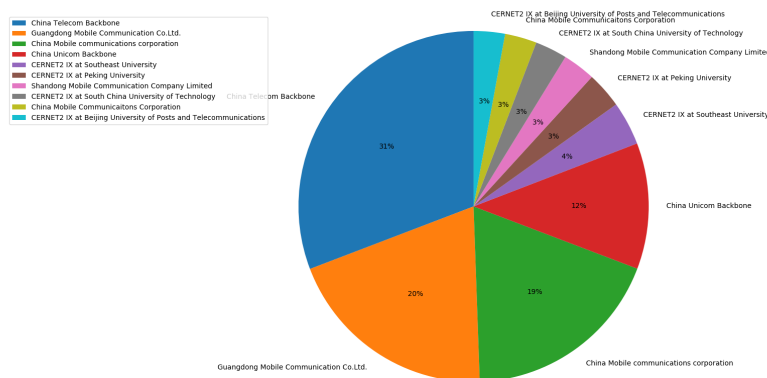


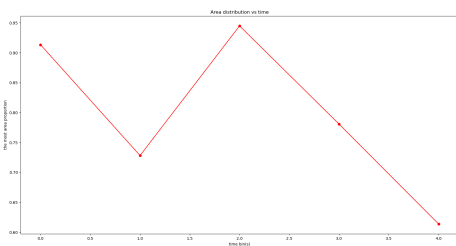
Figure 9: 最多的 10 个运营商信息饼状图

4 运营商、所在地域的时间分布分析

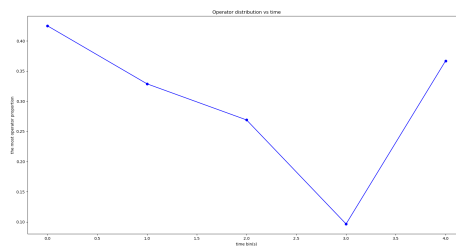
首先我们根据 BGP 路由表中的时间划分时隙，在本次大作业中，我们采用一周作为一个时隙，来研究上述两个分布随着时间的变化。根据 BGP 路由表的时间信息，共有 5 个时间段。

4.1 IPv4 路由表运营商、所在地域的时间分布分析

首先我们绘制出最多的地域和最多的运营商占总数量的比例随着时间的变化如图10所示，可以看出以一周为时间间隔的话，最多的地域和最多的运行商都是随机变化的。然后，类似于上一章的算法，我们分别绘制出 5 个时间段的最多的 10 个地域信息饼状图以及最多的 10 个运营商信息饼状图如图12所示，可以看出，虽然最多的地域的比例在不断的变化，但是最多的地域一直都是“CN”中国，这也与我们的直观认识相符。但是随着时间的变化，运营商的比例以及占比最多的运营商都在发生变化。



(a) 最多的地域占比随时间的变化图

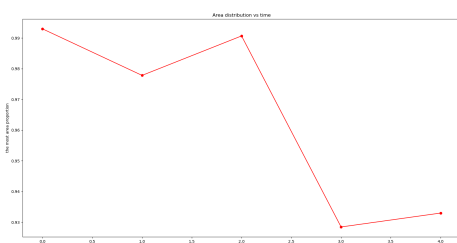


(b) 最多的运营商占比随时间的变化图

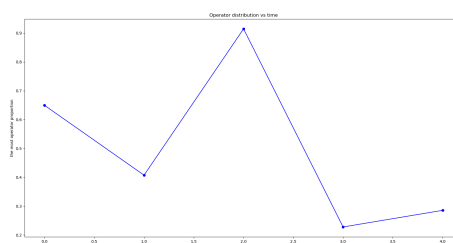
Figure 10: 最多的地域和最多的运营商占总数量的比例随着时间变化图

4.2 IPv6 路由表运营商、所在地域的时间分布分析

同样的道理，我们绘制出 IPv6 路由表中最多的地域和最多的运营商占总数量的比例随着时间的变化如图11所示，可以看出以一周为时间间隔的话，最多的地域和最多的运行商都是随机变化的。然后，类似于上一章的算法，我们分别绘制出 5 个时间段的最多的 10 个地域信息饼状图以及最多的 10 个运营商信息饼状图如图12所示，可以看出，虽然最多的地域的比例在不断的变化，但是最多的地域一直都是“CN”中国，并且 IPv6 路由表中，中国的占比更高。但是随着时间的变化，运营商的比例以及占比最多的运营商依然也都在发生变化。



(a) 最多的地域占比随时间的变化图

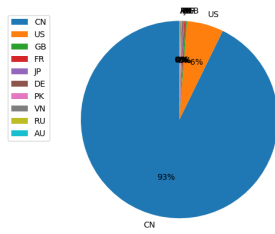


(b) 最多的运营商占比随时间的变化图

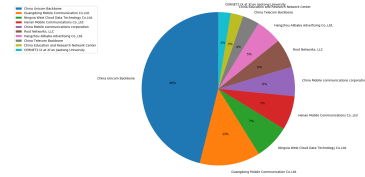
Figure 11: 最多的地域和最多的运营商占总数量的比例随着时间变化图

5 致谢

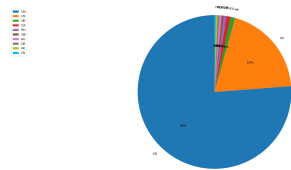
- 感谢李星老师与黄永峰老师一学期的辛勤授课；
- 感谢助教一学期辛苦批改作业以及对大作业的答疑。



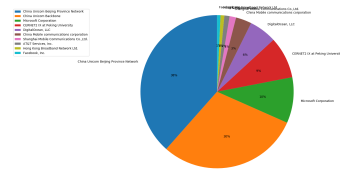
(a) 第一周地域分布图



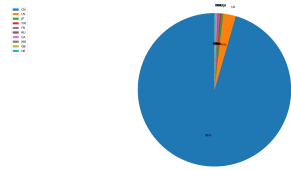
(b) 第一周运营商分布图



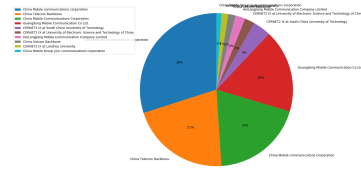
(c) 第二周地域分布图



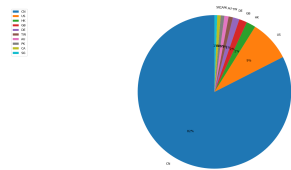
(d) 第二周运营商分布图



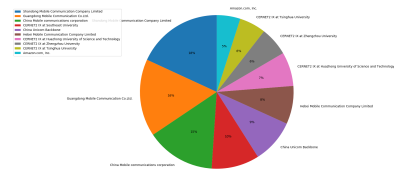
(e) 第三周地域分布图



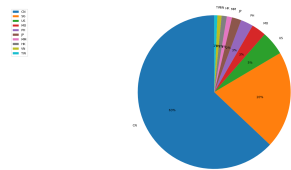
(f) 第三周运营商分布图



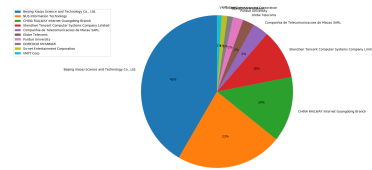
(g) 第四周地域分布图



(h) 第四周运营商分布图



(i) 第五周地域分布图



(j) 第五周运营商分布图

Figure 12: IPv4 路由表各周 AS 所在地域以及运营商分布饼状图

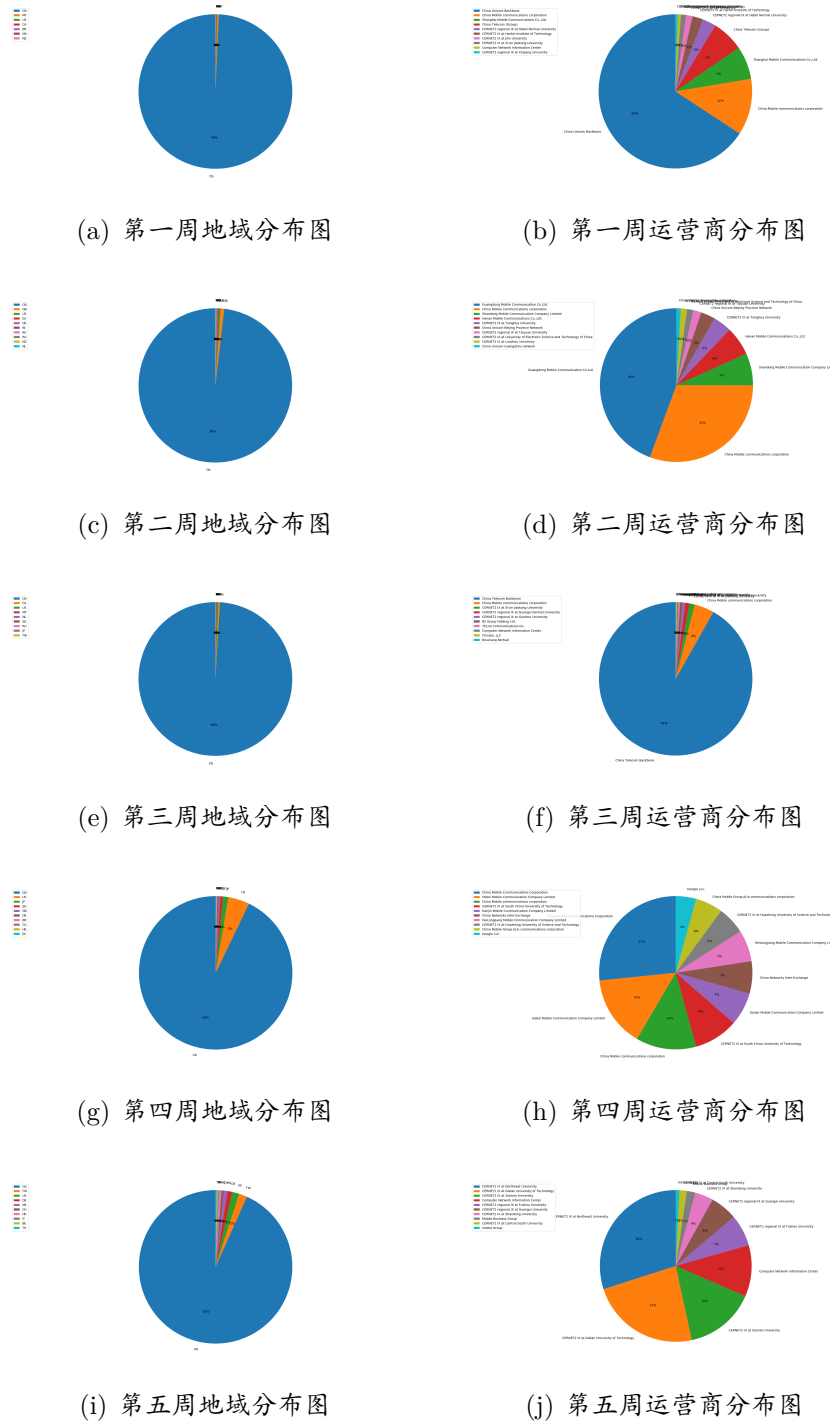


Figure 13: IPv6 路由表每周 AS 所在地域以及运营商分布饼状图