2016 年中国可视化与可视分析大会

数据可视分析挑战赛-挑战 1

（ChinaVis Data Challenge 2016 - mini challenge 1）

答　卷

参赛队名称：　　燕山大学-栾鑫月

团队成员：　栾鑫月，燕山大学，464176845@qq.com，队长

王美，燕山大学，1244431884@qq.com

聂俊岚，燕山大学，niejl3@163.com，指导老师

是否学生队（是或否）：　是

使用的分析工具或开发工具：D3，MySQL，Gephi，Tableau

共计耗费时间（人天）：　60 人天

本次比赛结束后，我们是否可以在网络上公布该答卷与视频（是或否）：是

挑战 1.1： 找出 BigBusiness 公司内部网络中的客户端与服务器，并给出 BigBusiness 公司的网络体系结构拓扑图；对 BigBusiness 内部网络中的服务器进行分类，分类标准不限，比如:按照节点类型、按时间特点、按行为特点、按流量特点等等。（请将回答尽量限制在 1500 个字和 10 张图片内）

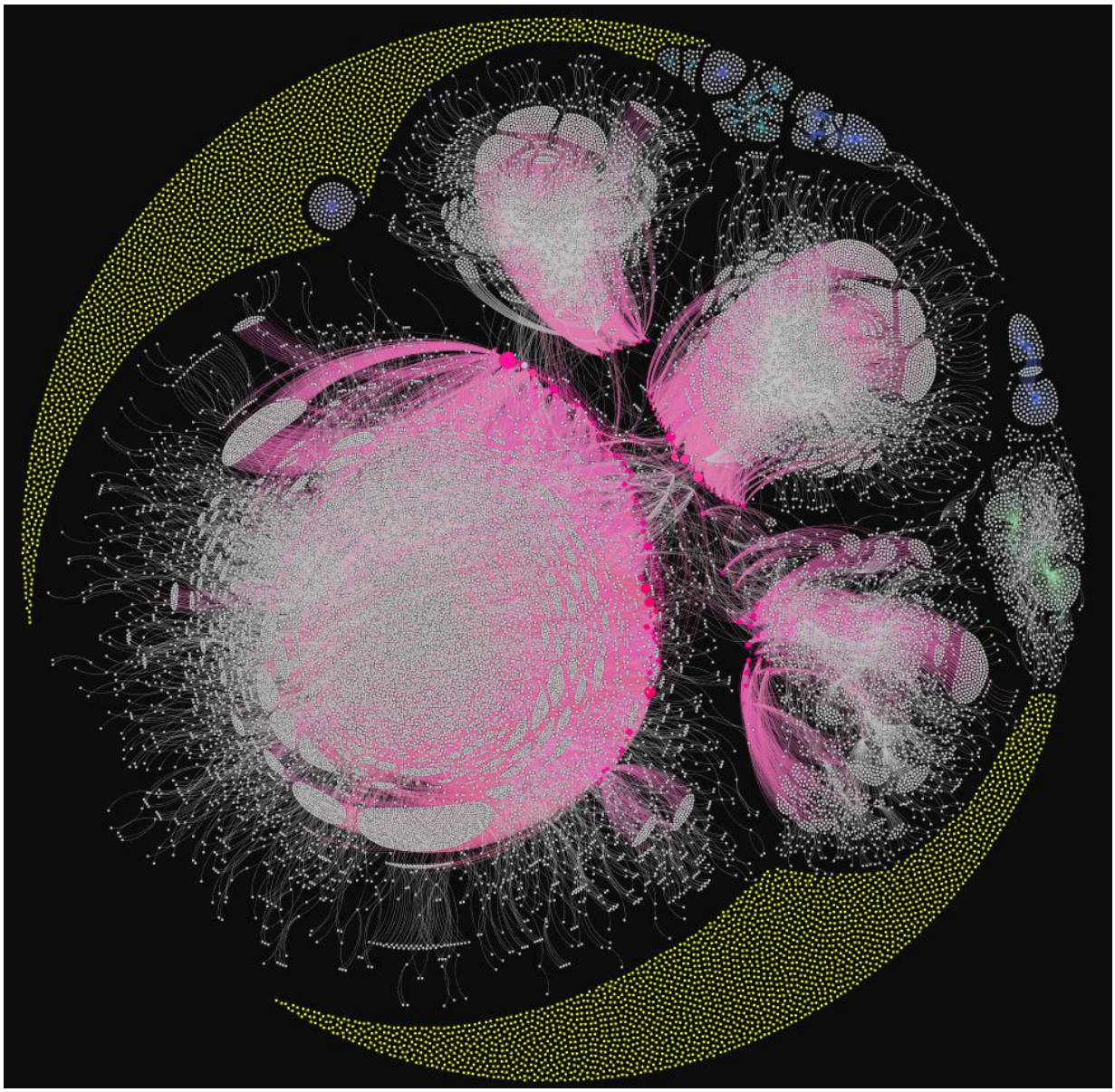

图 1-1-1 网络连接图

图 1-1-1 表示给出的数据中所有主机及主机间的联系，粉色节点表示找到的服务器，粉色的线表示服务器的通信路径。进一步观察上图，可将主机分为如下四部分：

1）黄色的节点：黄色节点代表的主机是在网络中关系简单（只与十个以内的主机通信，大部分只与一个主机通信）的主机，这些节点与内网主要网络无关联，不是分析的主要对象。

2）粉色的节点：粉色节点代表的是找到的服务器，这部分节点构成的网络基本上连接了内网的所有客户端（白色节点），且活跃周期较长，因此，认为粉色的服务器是正确的。

3）蓝色的节点：蓝色节点代表的主机也是找到的服务器，但这部分服务器存在的时间很短，且与这些服务器联系的节点所形成的网络与粉色节点所在的主要网络不存在联系，因此在服务器中排除这些节点，详细分析见挑战1.2。

4）绿色的节点：绿色的节点和蓝色节点的区别在于活跃的时间不一样，在服务器中也排除这些节点，详细分析见挑战 1.2。

综上，找到的服务器如下：

表 1-1-1 内部网络服务器 IP 地址

编号	IP	编号	IP	编号	IP	编号	IP
1	10.114.216.221	29	10.18.112.247	57	10.24.249.221	85	10.52.140.240
2	10.114.218.221	30	10.18.112.254	58	10.24.249.225	86	10.52.140.74
3	10.114.219.221	31	10.18.112.26	59	10.24.249.227	87	10.52.140.75
4	10.116.160.247	32	10.18.112.30	60	10.24.249.228	88	10.52.171.245
5	10.116.160.248	33	10.18.113.246	61	10.24.88.206	89	10.52.176.250
6	10.116.217.221	34	10.18.114.246	62	10.24.88.207	90	10.52.176.3
7	10.118.128.206	35	10.18.116.246	63	10.25.88.206	91	10.52.206.132
8	10.118.161.1	36	10.18.120.246	64	10.26.112.246	92	10.52.63.80
9	10.118.161.2	37	10.18.240.246	65	10.26.88.206	93	10.54.0.206
10	10.118.165.199	38	10.18.48.246	66	10.28.88.206	94	10.54.128.206
11	10.118.216.221	39	10.18.80.246	67	10.33.216.221	95	10.54.128.207
12	10.144.216.221	40	10.18.96.246	68	10.33.219.246	96	10.54.160.206
13	10.145.216.221	41	10.182.128.206	69	10.37.216.221	97	10.54.192.206
14	10.146.112.246	42	10.19.112.246	70	10.38.128.206	98	10.54.216.221
15	10.146.216.221	43	10.2.112.246	71	10.39.216.247	99	10.55.128.206
16	10.147.216.221	44	10.20.1.207	72	10.46.236.221	100	10.62.128.206
17	10.152.88.206	45	10.20.1.31	73	10.46.48.1	101	10.65.220.221
18	10.16.112.246	46	10.22.0.190	74	10.50.112.246	102	10.67.216.221
19	10.16.88.206	47	10.22.112.246	75	10.50.128.206	103	10.67.216.226
20	10.18.112.118	48	10.22.128.206	76	10.52.128.206	104	10.67.220.221
21	10.18.112.182	49	10.24.120.206	77	10.52.138.191	105	10.67.220.225
22	10.18.112.214	50	10.24.124.230	78	10.52.138.193	106	10.67.220.248
23	10.18.112.222	51	10.24.157.225	79	10.52.138.21	107	10.8.88.206
24	10.18.112.230	52	10.24.158.230	80	10.52.138.3	108	10.82.112.246
25	10.18.112.241	53	10.24.216.206	81	10.52.140.107	109	10.88.88.206
26	10.18.112.242	54	10.24.24.206	82	10.52.140.13	110	10.65.216.226
27	10.18.112.244	55	10.24.248.242	83	10.52.140.210	111	10.52.140.228
28	10.18.112.246	56	10.24.248.253	84	10.52.140.211		

网络结构如下：

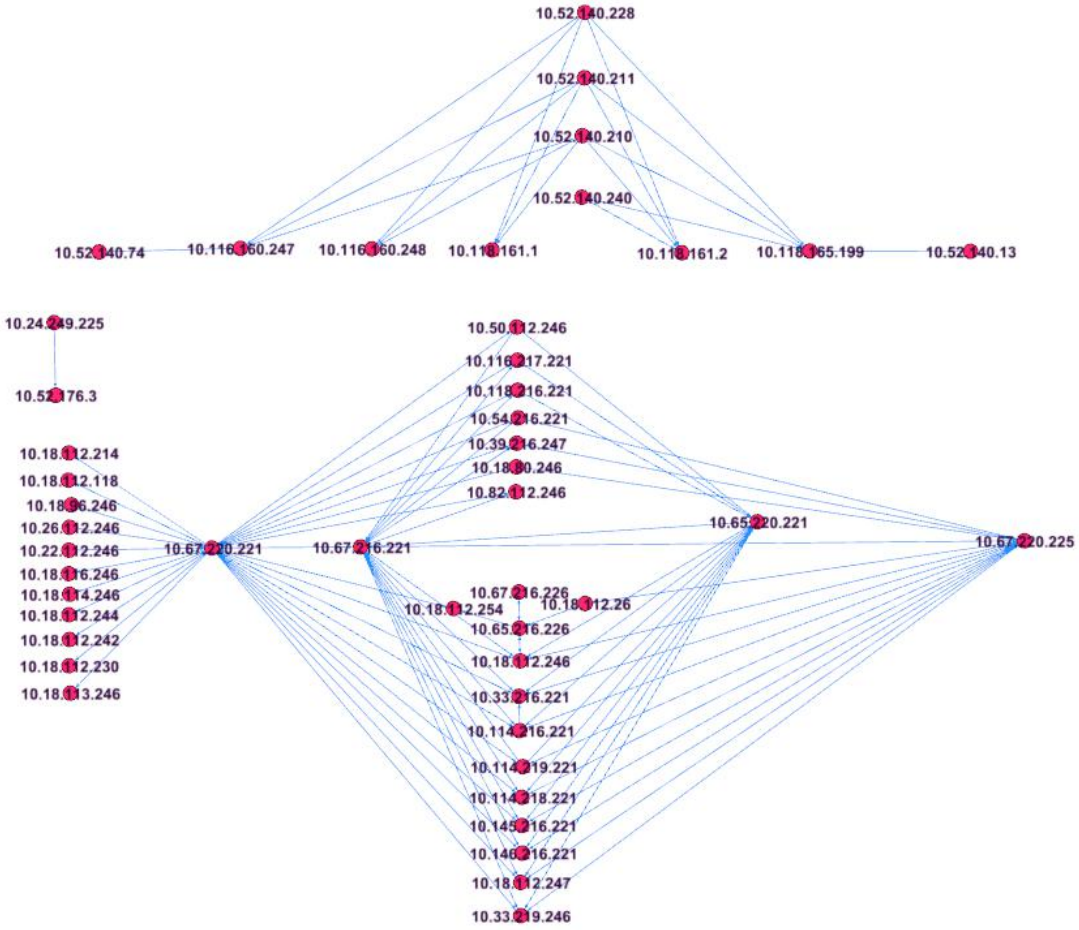

图 1-1-2 网络结构图

问题 2：服务器分类

1）按照服务器提供的服务分类：

一台服务器可能会使用多个端口，但每个端口的使用次数占该服务器总使用次数的比重不同，以使用比重最大的端口最为判定该服务器提供服务类型的依据，具体分类如下：

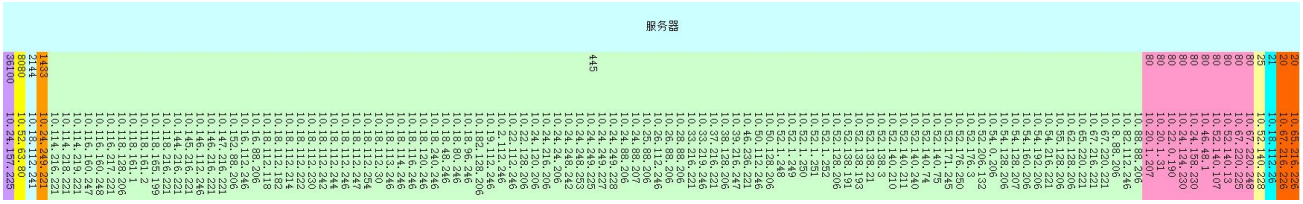

图 1-2-1 服务器应用分类图

2）按照入度和出度大小是否一致分类：

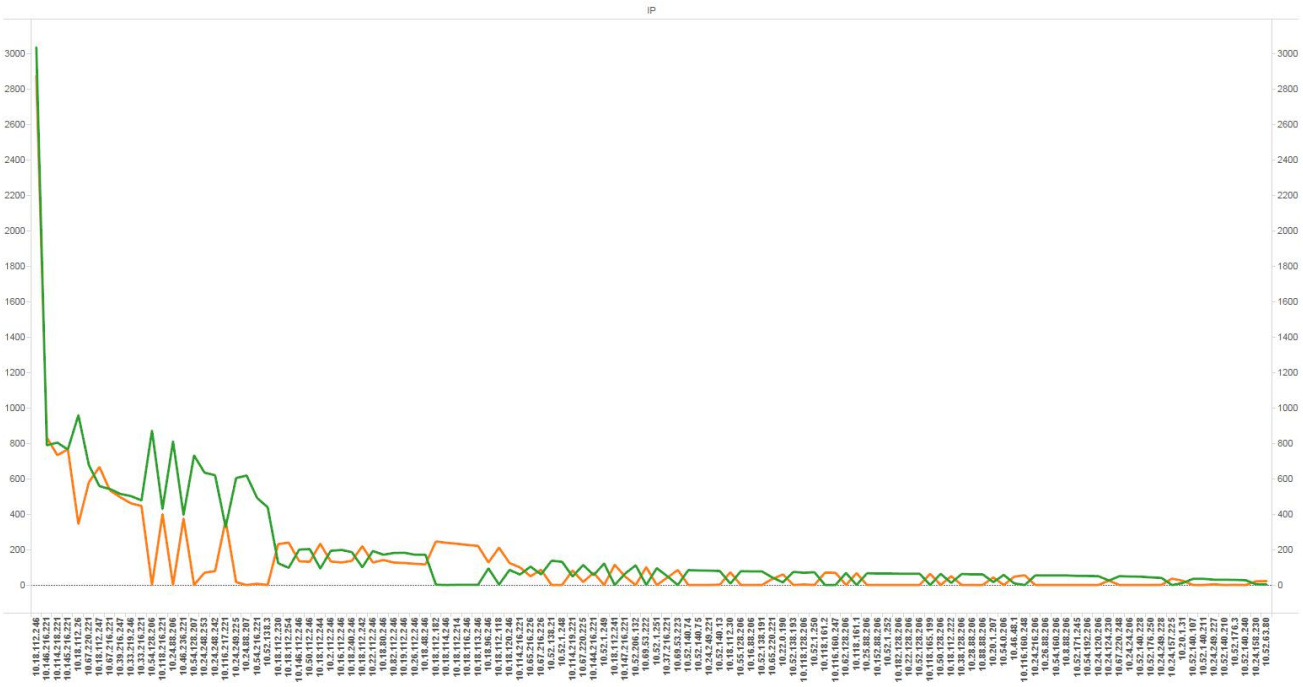

图 1-2-2 服务器入度出度关系图

上图中绿色和橘黄色的折线分别表示服务器的入度和出度，由图可知，有的服务器入度和出度相差不大（两条线基本重合），有的差别较大，观察上图，按照入度和出度的差在 50 以内为入度出度一致，否则不一致为标准，将服务器分为以下两类：

表 1-2-1 服务器入度出度是否一致分类表

出度与入度是否一致	IP
否	10. 114. 216. 221, 10. 116. 217. 221, 10. 118. 216. 221, 10. 144. 216. 221, 10. 145. 216. 221, 10. 146. 216. 221, 10. 147. 216. 221, 10. 18. 112. 222, 10. 18. 120. 246, 10. 18. 240. 246, 10. 18. 96. 246, 10. 20. 1. 207, 10. 24. 124. 230, 10. 24. 157. 225, 10. 24. 158. 230, 10. 24. 249. 228, 10. 33. 216. 221, 10. 33. 219. 246, 10. 37. 216. 221, 10. 39. 216. 247, 10. 46. 236. 221, 10. 46. 48. 1, 10. 52. 140. 107, 10. 52. 63. 80, 10. 65. 220. 221, 10. 67. 216. 221, 10. 114. 219. 221, 10. 67. 216. 226
是	10. 114. 218. 221, 10. 116. 160. 247, 10. 116. 160. 248, 10. 118. 128. 206, 10. 118. 161. 1, 10. 118. 161. 2, 10. 118. 165. 199, 10. 146. 112. 246, 10. 152. 88. 206, 10. 16. 112. 246, 10. 16. 88. 206, 10. 18. 112. 118, 10. 18. 112. 182, 10. 18. 112. 214, 10. 18. 112. 230, 10. 18. 112. 241, 10. 18. 112. 242, 10. 18. 112. 244, 10. 18. 112. 246, 10. 18. 112. 247, 10. 18. 112. 254, 10. 18. 112. 26, 10. 18. 112. 30, 10. 18. 113. 246, 10. 18. 114. 246, 10. 18. 116. 246, 10. 18. 48. 246, 10. 18. 80. 246, 10. 182. 128. 206, 10. 19. 112. 246, 10. 2. 112. 246, 10. 20. 1. 31, 10. 22. 0. 190, 10. 22. 112. 246, 10. 22. 128. 206, 10. 24. 120. 206, 10. 24. 216. 206, 10. 24. 24. 206, 10. 24. 248. 242, 10. 24. 248. 253, 10. 24. 249. 221, 10. 24. 249. 225, 10. 24. 249. 227, 10. 24. 88. 206, 10. 24. 88. 207, 10. 25. 88. 206, 10. 26. 112. 246, 10. 26. 88. 206, 10. 28. 88. 206, 10. 38. 128. 206, 10. 50. 112. 246, 10. 50. 128. 206, 10. 52. 128. 206, 10. 52. 138. 191, 10. 52. 138. 193, 10. 52. 138. 21, 10. 52. 138. 3, 10. 52. 140. 13, 10. 52. 140. 210, 10. 52. 140. 211, 10. 52. 140. 228, 10. 52. 140. 240, 10. 52. 140. 74, 10. 52. 140. 75, 10. 52. 171. 245, 10. 52. 176. 250, 10. 52. 176. 3, 10. 52. 206. 132, 10. 54. 0. 206, 10. 54. 128. 206, 10. 54. 128. 207, 10. 54. 160. 206, 10. 54. 192. 206, 10. 54. 216. 221, 10. 55. 128. 206, 10. 62. 128. 206, 10. 65. 216. 226, 10. 67. 220. 221, 10. 67. 220. 225, 10. 67. 220. 248, 10. 8. 88. 206, 10. 82. 112. 246, 10. 88. 88. 206

3）按照发送和接收流量大小是否一致分类：

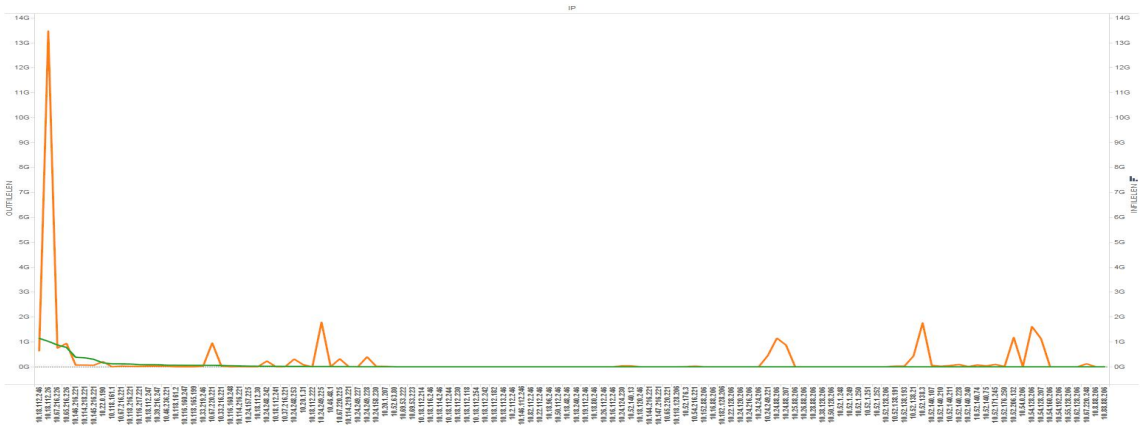

图 1-2-3 服务器接收和发出流量关系图

上图中绿色和黄色的曲线分别代表服务器接收和发送的流量，由图可知，大部分服务器接收和发送的流量都较少，少数的服务器接收流量少但发送的流量却很多，因此，按照接收和发送流量水平是否一致将服务器分为以下两类：

表 1-2-2 服务器接收和发送流量是否一致分类表

接收与发送流量是否一致	IP
否	10. 114. 218. 221, 10. 145. 216. 221, 10. 146. 216. 221, 10. 18. 112. 246, 10. 18. 112. 26, 10. 24. 248. 242, 10. 24. 248. 253, 10. 24. 249. 221, 10. 24. 249. 225, 10. 24. 249. 228, 10. 24. 88. 207, 10. 25. 88. 206, 10. 52. 138. 3, 10. 52. 206. 132, 10. 54. 128. 206, 10. 54. 128. 207, 10. 67. 220. 221, 10. 67. 220. 225, 10. 67. 220. 248
是	10. 114. 216. 221, 10. 114. 219. 221, 10. 116. 160. 247, 10. 116. 160. 248, 10. 116. 217. 221, 10. 118. 128. 206, 10. 118. 161. 1, 10. 118. 161. 2, 10. 118. 165. 199, 10. 118. 216. 221, 10. 144. 216. 221, 10. 146. 112. 246, 10. 147. 216. 221, 10. 152. 88. 206, 10. 16. 112. 246, 10. 16. 88. 206, 10. 18. 112. 118, 10. 18. 112. 182, 10. 18. 112. 214, 10. 18. 112. 222, 10. 18. 112. 230, 10. 18. 112. 241, 10. 18. 112. 242, 10. 18. 112. 244, 10. 18. 112. 247, 10. 18. 112. 254, 10. 18. 112. 30, 10. 18. 113. 246, 10. 18. 114. 246, 10. 18. 116. 246, 10. 18. 120. 246, 10. 18. 240. 246, 10. 18. 48. 246, 10. 18. 80. 246, 10. 18. 96. 246, 10. 182. 128. 206, 10. 19. 112. 246, 10. 2. 112. 246, 10. 20. 1. 207, 10. 20. 1. 31, 10. 22. 0. 190, 10. 22. 112. 246, 10. 22. 128. 206, 10. 24. 120. 206, 10. 24. 124. 230, 10. 24. 157. 225, 10. 24. 158. 230, 10. 24. 216. 206, 10. 24. 24. 206, 10. 24. 249. 227, 10. 24. 88. 206, 10. 26. 112. 246, 10. 26. 88. 206, 10. 28. 88. 206, 10. 33. 216. 221, 10. 33. 219. 246, 10. 37. 216. 221, 10. 38. 128. 206, 10. 39. 216. 247, 10. 46. 236. 221, 10. 46. 48. 1, 10. 50. 112. 246, 10. 50. 128. 206, 10. 52. 128. 206, 10. 52. 138. 191, 10. 52. 138. 193, 10. 52. 138. 21, 10. 52. 140. 107, 10. 52. 140. 13, 10. 52. 140. 210, 10. 52. 140. 211, 10. 52. 140. 228, 10. 52. 140. 240, 10. 52. 140. 74, 10. 52. 140. 75, 10. 52. 171. 245, 10. 52. 176. 250, 10. 52. 176. 3, 10. 52. 63. 80, 10. 54. 0. 206, 10. 54. 160. 206, 10. 54. 192. 206, 10. 54. 216. 221, 10. 55. 128. 206, 10. 62. 128. 206, 10. 65. 216. 226, 10. 65. 220. 221, 10. 67. 216. 221, 10. 67. 216. 226, 10. 8. 88. 206, 10. 82. 112. 246, 10. 88. 88. 206

4）按照传输流量大小分类：

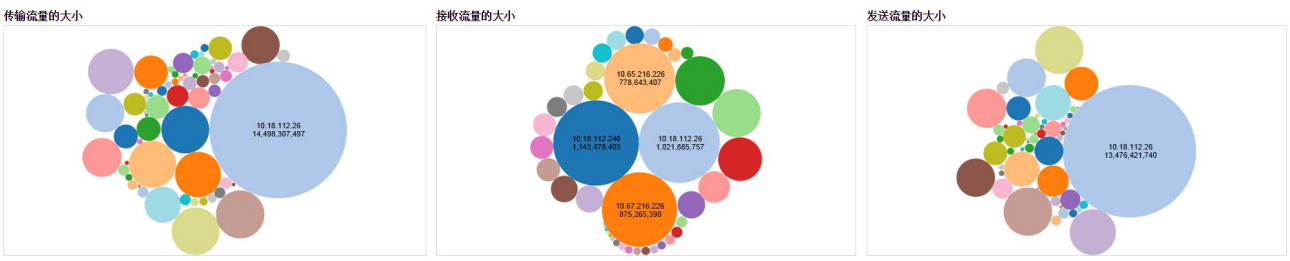

图 1-2-4 服务器传输流量大小气泡图

上图中不同颜色的气泡代表不同的服务器，气泡的大小代表服务器传输流量的大小，左边的气泡图代表服务器传输的总流量大小，中间的气泡图代表服务器接收流量的大小，右边的气泡图代表服务器发送流量的大小。按照传输的总流量、接收和发送流量的大小分类，具体分类情况如表 1-2-3 所示：

表 1-2-3 服务器传输流量分类表

传输数据量是否很大	是	否
发送流量	10. 18. 112. 26, 10. 24. 249. 225, 10. 52. 138. 3, 10. 54. 128. 206	10. 52. 206. 132, 10. 24. 88. 206, 10. 54. 128. 207, 10. 67. 220. 221, 10. 65. 216. 226, 10. 24. 88. 207, 10. 67. 216. 226, 10. 18. 112. 246, 10. 24. 249. 221, 10. 52. 138. 21, 10. 24. 249. 228, 10. 67. 220. 225, 10. 24. 248. 253, 10. 24. 248. 242, 10. 22. 0. 190, 10. 67. 220. 248, 10. 52. 171. 245, 10. 52. 140. 228, 10. 20. 1. 31, 10. 52. 140. 74, 10. 146. 216. 221, 10. 114. 218. 221, 10. 145. 216. 221, 10. 52. 140. 107, 10. 52. 140. 211, 10. 24. 124. 230, 10. 52. 140. 75, 10. 52. 140. 13, 10. 67. 216. 221, 10. 52. 138. 191, 10. 52. 138. 193, 10. 54. 216. 221, 10. 33. 216. 221, 10. 18. 112. 247, 10. 39. 216. 247, 10. 33. 219. 246, 10. 52. 140. 210, 10. 46. 236. 221, 10. 118. 216. 221, 10. 24. 158. 230, 10. 20. 1. 207, 10. 52. 176. 250, 10. 116. 217. 221, 10. 52. 140. 240, 10. 24. 249. 227, 10. 52. 176. 3, 10. 114. 216. 221, 10. 114. 219. 221, 10. 46. 48. 1, 10. 37. 216. 221, 10. 38. 128. 206, 10. 18. 112. 222, 10. 54. 160. 206, 10. 50. 128. 206, 10. 152. 88. 206, 10. 182. 128. 206, 10. 88. 88. 206, 10. 52. 128. 206, 10. 55. 128. 206, 10. 62. 128. 206, 10. 54. 0. 206, 10. 54. 192. 206, 10. 26. 88. 206, 10. 118. 128. 206, 10. 16. 88. 206, 10. 22. 128. 206, 10. 24. 24. 206, 10. 22. 112. 246, 10. 8. 88. 206, 10. 16. 112. 246, 10. 19. 112. 246, 10. 28. 88. 206, 10. 24. 120. 206, 10. 2. 112. 246, 10. 24. 216. 206, 10. 25. 88. 206, 10. 146. 112. 246, 10. 18. 80. 246, 10. 18. 48. 246, 10. 50. 112. 246, 10. 18. 240. 246, 10. 26. 112. 246, 10. 82. 112. 246, 10. 52. 63. 80, 10. 65. 220. 221, 10. 18. 112. 230, 10. 18. 112. 242, 10. 18. 112. 244, 10. 18. 96. 246, 10. 18. 120. 246, 10. 18. 112. 254, 10. 144. 216. 221, 10. 147. 216. 221, 10. 18. 112. 30, 10. 18. 112. 182, 10. 18. 112. 118, 10. 18. 113. 246, 10. 18. 116. 246, 10. 18. 112. 21410. 116. 160. 247, 10. 116. 160. 248, 10. 118. 161. 1, 10. 118. 161. 2, 10. 118. 165. 199, 10. 18. 112. 241, 10. 18. 114. 246, 10. 24. 157. 225
接收流量	10. 18. 112. 246, 10. 18. 112. 26, 10. 67. 216. 226, 10. 65. 216. 226, 10. 146. 216. 221, 10. 114. 218. 221, 10. 145. 216. 221	10. 22. 0. 190, 10. 118. 161. 1, 10. 67. 216. 221, 10. 118. 216. 221, 10. 116. 217. 221, 10. 18. 112. 247, 10. 39. 216. 247, 10. 46. 236. 221, 10. 118. 161. 2, 10. 116. 160. 247, 10. 118. 165. 199, 10. 33. 219. 246, 10. 67. 220. 221, 10. 33. 216. 221, 10. 116. 160. 248, 10. 114. 216. 221, 10. 24. 157. 225, 10. 18. 112. 30, 10. 24. 248. 242, 10. 18. 112. 241, 10. 37. 216. 221, 10. 24. 248. 253, 10. 20. 1. 31, 10. 18. 112. 222, 10. 24. 249. 225, 10. 46. 48. 1, 10. 67. 220. 225, 10. 114. 219. 221, 10. 24. 249. 227, 10. 24. 249. 228, 10. 24. 158. 230, 10. 20. 1. 207, 10. 52. 63. 80, 10. 18. 112. 214, 10. 18. 116. 246, 10. 18. 114. 246, 10. 18. 112. 244, 10. 18. 112. 230, 10. 18. 112. 118, 10. 18. 112. 254, 10. 18. 112. 242, 10. 18. 112. 182, 10. 18. 113. 246, 10. 2. 112. 246, 10. 146. 112. 246, 10. 82. 112. 246, 10. 22. 112. 246, 10. 18. 96. 246, 10. 50. 112. 246, 10. 18. 48. 246, 10. 18. 240. 246, 10. 19. 112. 246, 10. 18. 80. 246, 10. 26. 112. 246, 10. 16. 112. 246, 10. 24. 124. 230, 10. 52. 140. 13, 10. 18. 120. 246, 10. 144. 216. 221, 10. 147. 216. 221, 10. 65. 220. 221, 10. 118. 128. 206, 10. 52. 176. 3, 10. 54. 216. 221, 10. 152. 88. 206, 10. 16. 88. 206, 10. 182. 128. 206, 10. 22. 128. 206, 10. 24. 120. 206, 10. 24. 216. 206, 10. 24. 24. 206, 10. 24. 249. 221, 10. 24. 88. 206, 10. 24. 88. 207, 10. 25. 88. 206, 10. 26. 88. 206, 10. 28. 88. 206, 10. 38. 128. 206, 10. 50. 128. 206, 10. 52. 128. 206, 10. 52. 138. 191, 10. 52. 138. 193, 10. 52. 138. 21, 10. 52. 138. 3, 10. 52. 140. 107, 10. 52. 140. 210, 10. 52. 140. 211, 10. 52. 140. 240, 10. 52. 140. 74, 10. 52. 140. 75, 10. 52. 171. 245, 10. 52. 176. 250, 10. 52. 206. 132, 10. 54. 0. 206, 10. 54. 128. 206, 10. 54. 128. 207, 10. 54. 160. 206, 10. 54. 192. 206, 10. 55. 128. 206, 10. 62. 128. 206, 10. 67. 220. 248, 10. 8. 88. 206, 10. 88. 88. 206, 10. 52. 140. 228
总传输流量	10. 18. 112. 26, 10. 24. 249. 225, 10. 18. 112. 246, 10. 52. 138. 3, 10. 65. 216. 226, 10. 67. 216. 226, 10. 54. 128. 206,	10. 52. 206. 132, 10. 24. 88. 206, 10. 54. 128. 207, 10. 67. 220. 221, 10. 24. 88. 207, 10. 146. 216. 221, 10. 24. 249. 221, 10. 52. 138. 21, 10. 114. 218. 221, 10. 24. 249. 228, 10. 22. 0. 190, 10. 145. 216. 221, 10. 67. 220. 225, 10. 24. 248. 253, 10. 24. 248. 242, 10. 67. 216. 221, 10. 118. 216. 221, 10. 67. 220. 248, 10. 118. 161. 1, 10. 18. 112. 247, 10. 39. 216. 247, 10. 116. 217. 221, 10. 52. 171. 245, 10. 20. 1. 31, 10. 52. 140. 228, 10. 46. 236. 221, 10. 33. 219. 246, 10. 33. 216. 221, 10. 52. 140. 74, 10. 118. 161. 2, 10. 116. 160. 247, 10. 118. 165. 199, 10. 52. 140. 107, 10. 52. 140. 211, 10. 116. 160. 248, 10. 114. 216. 221, 10. 24. 124. 230, 10. 52. 140. 75, 10. 52. 140. 13, 10. 52. 138. 191, 10. 52. 138. 193, 10. 24. 157. 225, 10. 54. 216. 221, 10. 52. 140. 210, 10. 18. 112. 30, 10. 24. 158. 230, 10. 18. 112. 241, 10. 20. 1. 207, 10. 37. 216. 221, 10. 46. 48. 1, 10. 52. 176. 250, 10. 114. 219. 221, 10. 18. 112. 222, 10. 24. 249. 227, 10. 52. 140. 240, 10. 52. 176. 3, 10. 52. 63. 80, 10. 18. 112. 230, 10. 22. 112. 246, 10. 2. 112. 246, 10. 18. 112. 244, 10. 16. 112. 246, 10. 19. 112. 246, 10. 146. 112. 246, 10. 18. 112. 242, 10. 18. 112. 254, 10. 18. 48. 246, 10. 18. 80. 246, 10. 52. 1. 252, 10. 50. 112. 246, 10. 18. 240. 246, 10. 82. 112. 246, 10. 26. 112. 246, 10. 38. 128. 206, 10. 54. 160. 206, 10. 50. 128. 206, 10. 152. 88. 206, 10. 182. 128. 206, 10. 88. 88. 206, 10. 52. 128. 206, 10. 55. 128. 206, 10. 18. 112. 214, 10. 62. 128. 206, 10. 118. 128. 206, 10. 18. 116. 246, 10. 54. 0. 206, 10. 54. 192. 206, 10. 26. 88. 206, 10. 18. 114. 246, 10. 18. 112. 118, 10. 16. 88. 206, 10. 22. 128. 206, 10. 18. 96. 246, 10. 24. 24. 206, 10. 8. 88. 206, 10. 18. 112. 182, 10. 28. 88. 206, 10. 24. 120. 206, 10. 24. 216. 206, 10. 25. 88. 206, 10. 18. 113. 246, 10. 18. 120. 246, 10. 65. 220. 221, 10. 144. 216. 221, 10. 147. 216. 221

4）按照传输文件类型分类：

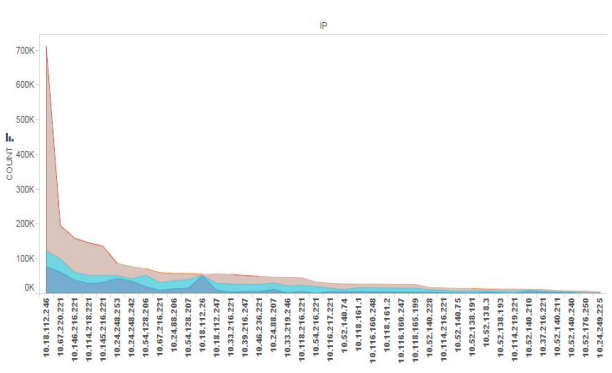

图 1-2-5 文件分类总图-河流图

图 1-2-6 文件分类详细-树地图

图 1-2-5 中从左至右服务器的通信次数依次降低，且传送的文件类型主要为空类型（棕黄色）、rpc 类型（天蓝色区域）、unk 类型（深蓝色区域）的文件；图 1-2-6 中，每个服务器的面积大小相同，对相同的面积进一步按照传输文件类型的比重进行划分，可以详细地看到每个服务器主要传输的文件类型，与河流图中看到的结果相同。按照主要传输的文件类型及活跃度对服务器分类如下：

表 1-2-4 服务器按照传输文件类型分类

主要传输文件类型	活跃度（传输次数）	IP
空类型、unk 和 rpc	高	10. 114. 216. 221, 10. 114. 218. 221, 10. 114. 219. 221, 10. 116. 160. 247, 10. 116. 160. 248, 10. 116. 217. 221, 10. 118. 161. 1, 10. 118. 161. 2, 10. 118. 165. 199, 10. 118. 216. 221, 10. 145. 216. 221, 10. 146. 216. 221, 10. 18. 112. 246, 10. 18. 112. 247, 10. 18. 112. 26, 10. 24. 248. 242, 10. 24. 248. 253, 10. 24. 249. 225, 10. 24. 88. 206, 10. 24. 88. 207, 10. 33. 216. 221, 10. 37. 216. 221, 10. 39. 216. 247, 10. 46. 236. 221, 10. 52. 138. 191, 10. 52. 138. 193, 10. 52. 138. 3, 10. 52. 140. 210, 10. 52. 140. 211, 10. 52. 140. 228, 10. 52. 140. 240, 10. 52. 140. 74, 10. 52. 140. 75, 10. 52. 176. 250, 10. 54. 128. 206, 10. 54. 128. 207, 10. 54. 216. 221, 10. 67. 216. 221, 10. 67. 220. 221, 10. 33. 219. 246
空类型或 unk 或 rpc	低	10. 118. 128. 206, 10. 144. 216. 221, 10. 146. 112. 246, 10. 147. 216. 221, 10. 152. 88. 206, 10. 16. 112. 246, 10. 16. 88. 206, 10. 18. 112. 118, 10. 18. 112. 182, 10. 18. 112. 214, 10. 18. 112. 222, 10. 18. 112. 230, 10. 18. 112. 241, 10. 18. 112. 242, 10. 18. 112. 244, 10. 18. 112. 254, 10. 18. 112. 30, 10. 18. 113. 246, 10. 18. 114. 246, 10. 18. 116. 246, 10. 18. 120. 246, 10. 18. 240. 246, 10. 18. 48. 246, 10. 18. 80. 246, 10. 18. 96. 246, 10. 182. 128. 206, 10. 19. 112. 246, 10. 2. 112. 246, 10. 20. 1. 207, 10. 20. 1. 31, 10. 22. 0. 190, 10. 22. 112. 246, 10. 22. 128. 206, 10. 24. 120. 206, 10. 24. 124. 230, 10. 24. 157. 225, 10. 24. 158. 230, 10. 24. 216. 206, 10. 24. 24. 206, 10. 24. 249. 221, 10. 24. 249. 227, 10. 24. 249. 228, 10. 25. 88. 206, 10. 26. 112. 246, 10. 26. 88. 206, 10. 28. 88. 206, 10. 38. 128. 206, 10. 46. 48. 1, 10. 50. 112. 246, 10. 50. 128. 206, 10. 52. 128. 206, 10. 52. 138. 21, 10. 52. 140. 107, 10. 52. 140. 13, 10. 52. 171. 245, 10. 52. 176. 3, 10. 52. 206. 132, 10. 52. 63. 80, 10. 54. 0. 206, 10. 54. 160. 206, 10. 54. 192. 206, 10. 55. 128. 206, 10. 62. 128. 206, 10. 65. 216. 226, 10. 65. 220. 221, 10. 67. 216. 226, 10. 67. 220. 225, 10. 67. 220. 248, 10. 8. 88. 206, 10. 82. 112. 246, 10. 88. 88. 206

5）按照活跃时段分类：

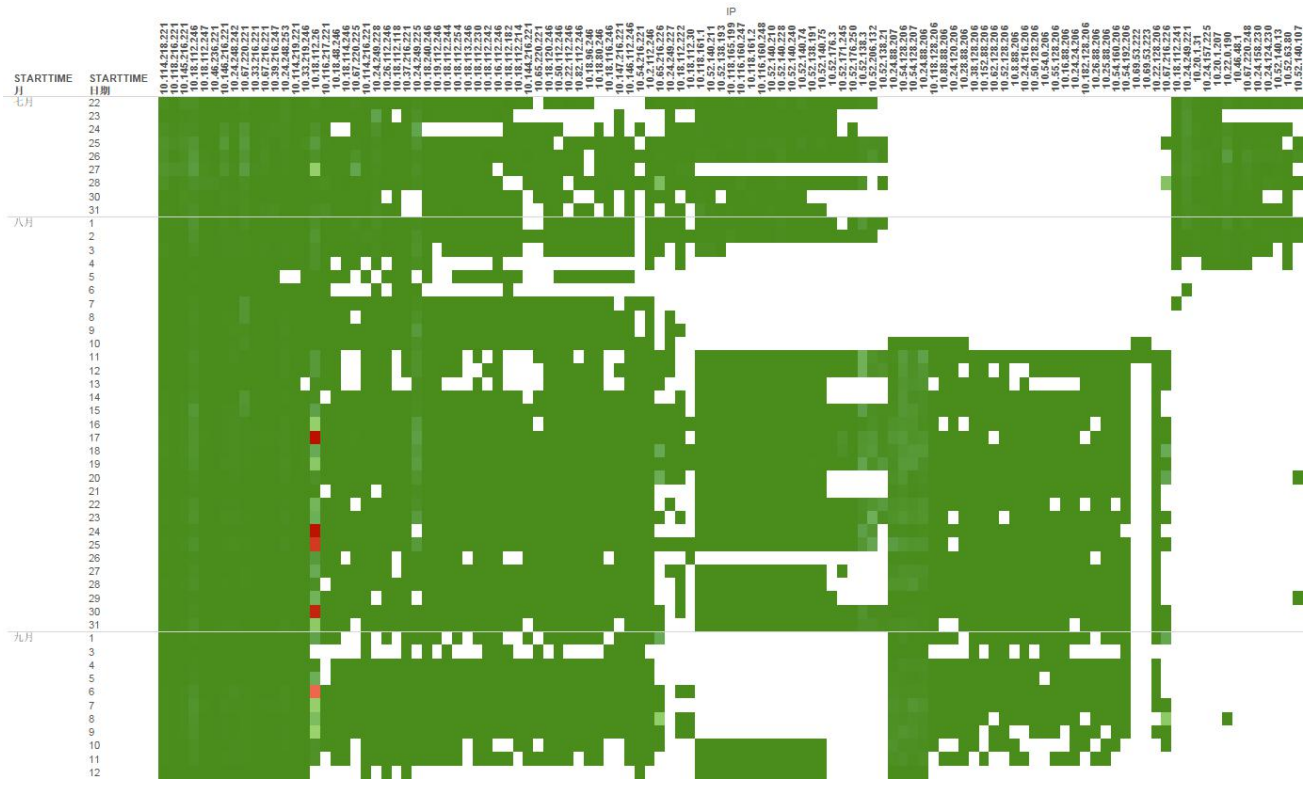

图 1-2-7 服务器活跃周期图

图 1-2-7 表示每个服务器在 53 天内的流量情况，红色越重表示流量越大，由图可知，服务器按照活跃的时段可以分为四类：在整个时间段内都活跃的（颜色填充是完整的），在整个时间段内基本上是活跃的（颜色填充有缺失但并不是很多），存在明显时间段内不活动情况的（8 月 3 日左右到 8 月 10 日左右和 9 月 1 日到 9 月 8 日左右颜色填充有缺失），只在 8 月 10 日后活跃的（8 月 10 日前无颜色填充），只在 8 月 4 日前活跃（8 月 4 日后颜色填充缺失）。具体分类情况如下图所示：

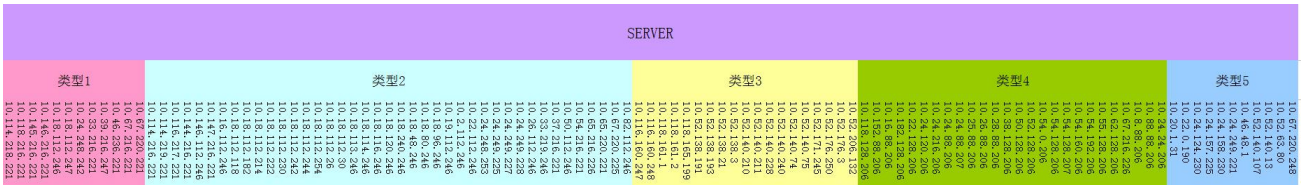

图 1-2-8 服务器按照活跃周期分类

类型 1：7 月 22 日-9 月 12 日，全部活跃

类型 2：7 月 22 日-9 月 12 日，有少量不活跃时段

类型 3：8 月 3 日-8 月 10 日和 9 月 1 日到 9 月 8 日不活跃

类型 4：只在 8 月 4 日前活跃

类型 5：只在 8 月 10 日后活跃

挑战 1.2： 找出可能存在的异常通信模式（异常事件），异常标准不限，比如：网络中的周期行为模式和活跃时段变化、某个子网访问或被访问量变化、子网连接模式变化、连接可能

传输的文件类型（exe、bmp、jpg 等）的变化等等，建议给出至少 5 种异常通信模式。（请将回答尽量限制在 1500 个字和 10 张图片内）

异常 1：网络结构异常

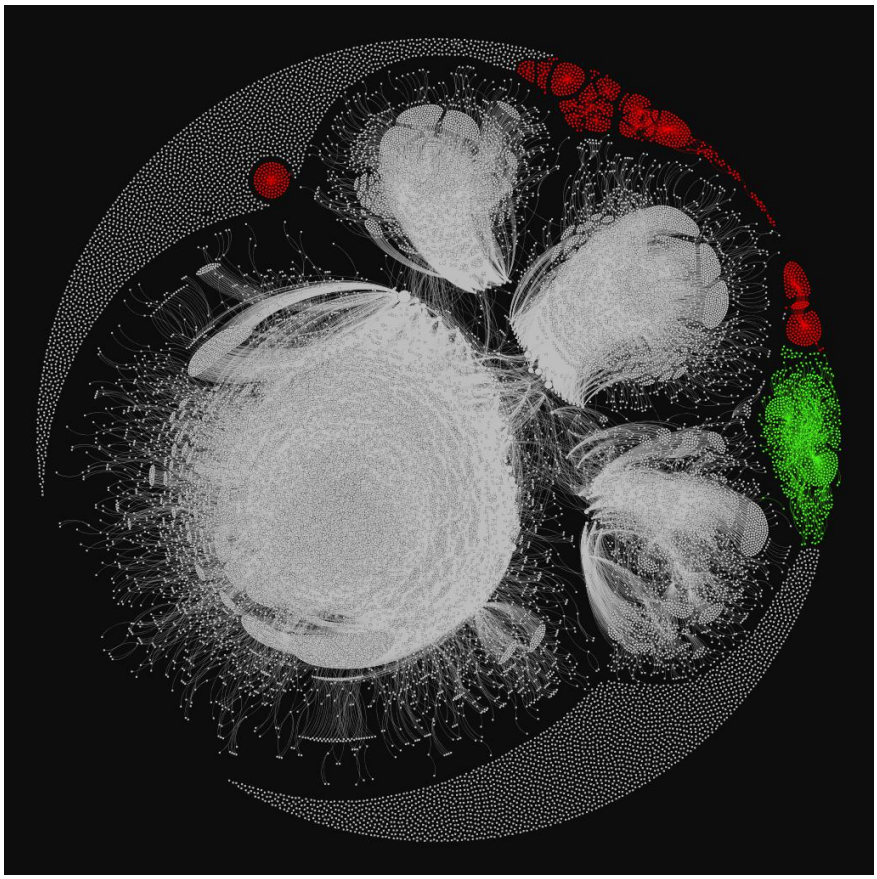

图 2-1-1 网络结构异常力引导图

上图中红色和绿色的主机不与主要网络通信，只在其内部通信，是孤立的，进一步分析发现如下异常：

上图中红色部分的IP都是只在2015/8/10 23:55:25 — 2015/8/11 0:31:58这半个小时内活跃，且以10.52.1.248，10.52.1.249，10.52.1.250，10.52.1.251，10.52.1.252，10.69.53.222，10.69.53.223，10.69.58.246，10.69.89.246，10.66.37.246，10.67.41.246，10.97.69.246，10.97.34.248为中心发生通信，与主干网络并没有通信。

上图中绿色部分的主机是以172.20.8.62和192.168.85.50为中心存在，主要与这两个主机通信，这些主机与红色部分主机的区别在于它们存在的时间比红色部分的主机长。172.20.8.62及与其通信的主机在2015/7/25 13:53:11 - 2015/7/31 9:23:26活跃，其它时间不活跃；192.168.85.50及与其通信的主机只在2015/7/25 12:43:00 - 2015/8/4 10:31:51内活跃。

异常 2：端口异常

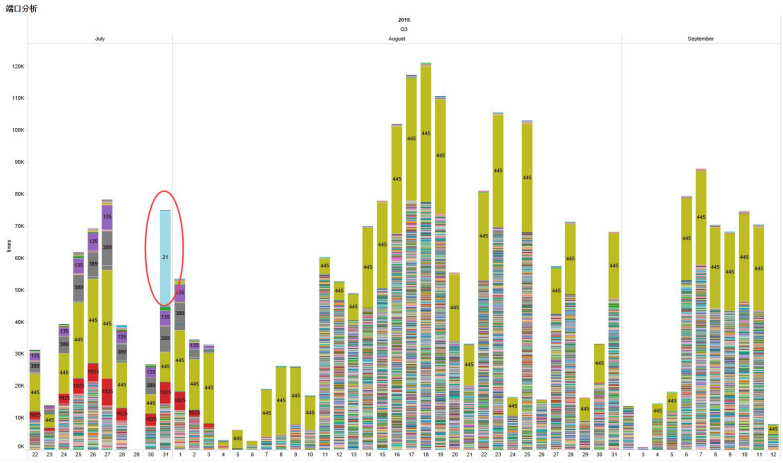

图 2-2-1 端口异常堆叠条

上图中，445端口（黄绿色柱形条）在整个时间内都是使用次数较多的；389端口（灰色柱形条）、135端口（紫色柱形条）和1025端口（红色柱形条）在8月3日前使用次数较多，此后使用次数较少； 21端口在7月31日被异常多次地使用，此前和此后使用频率较低。我们推测，389、135和1025端口的变化情况可能是由于企业业务调整使得应用服务不同最终造成了端口使用次数的变化，而21端口的异常，可能是遭到了FTP攻击。

异常 3：LAND attack

在2015/7/28 9:51:59到2015/7/28 10:03:00，2015/8/18 10:38:00和2015/9/8 5:56:49这三个时间点，也就是从7月28日开始的每隔三周的周二上午有主机以10.67.216.226为源和目的IP分别发送了27M、7M、1M左右的数据，怀疑是LAND attack，但该攻击并未成功，因为发生攻击后，该主机依然正常运行。

异常 4：通信次数异常

图2-4-1表示内网中通信次数随时间的变化，如图2-4-1左侧图所示，内网每天的通信次数在6000左右，7月31日，内网的通信次数突增，是正常通信次数的4倍左右。进一步分析发现10.65和10.18网段间的通信次数在当日有明显的异常。更进一步分析，发现两个网络内通信次数异常的主要原因是10.65.216.146向10.18.112.26发起了大量的通信造成的，如图2-4-1右侧部分所示。

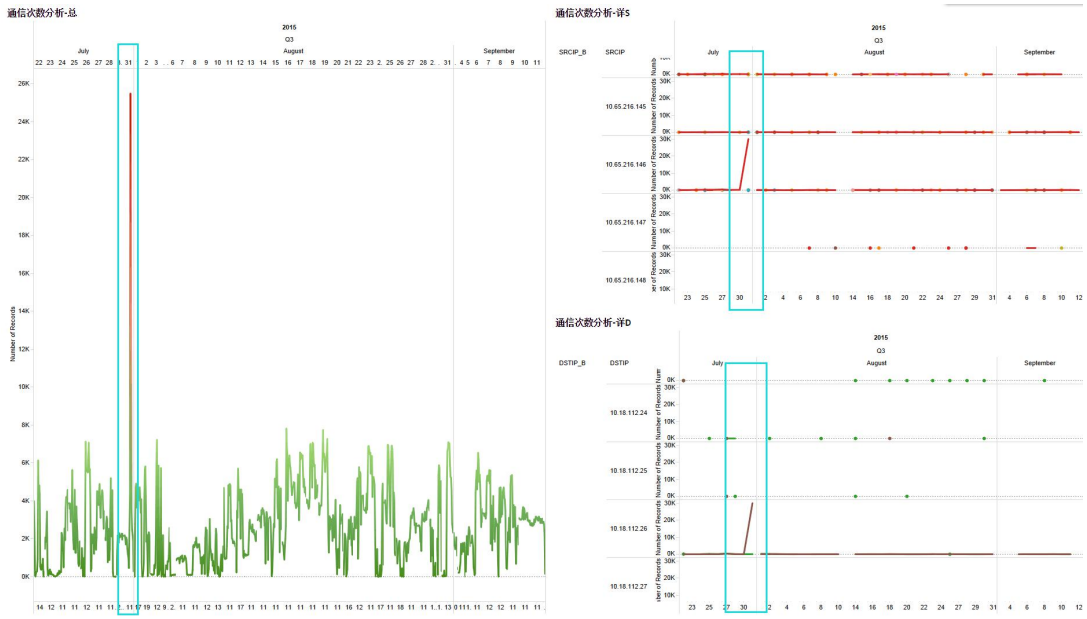

图 2-4-1 通信次数变化图

异常 5：传输流量异常

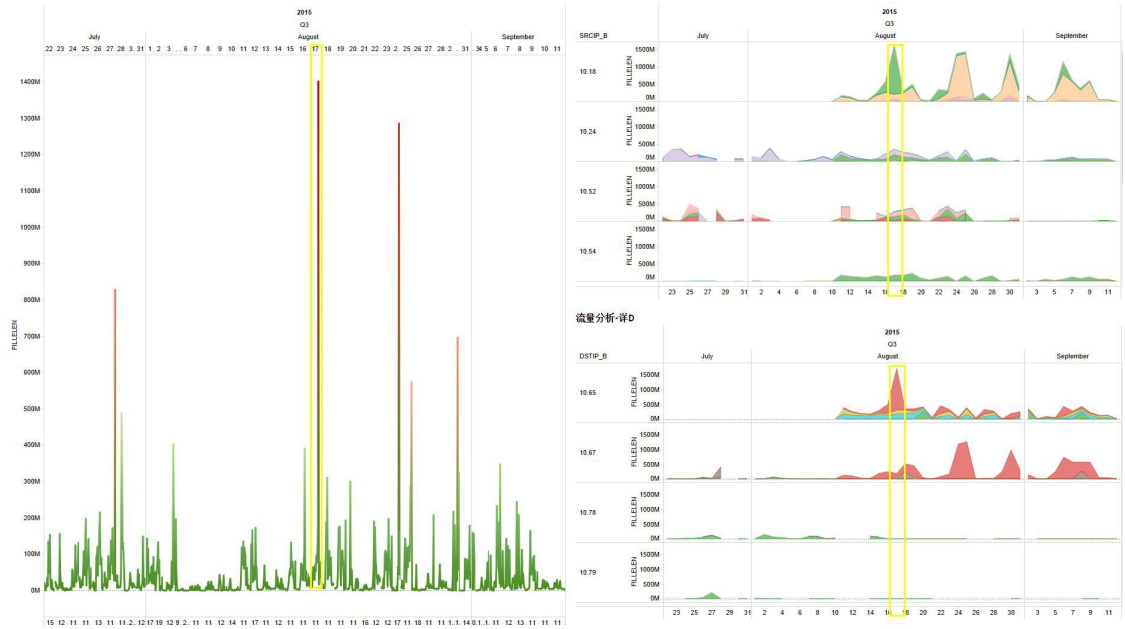

图 2-5-1 网段的流量变化图

图2-5-1表示内网中流量随时间的变化情况，在左侧流量分析-总图中可以看出在大部分时间每天总的传输流量在500M以下，左侧显示7月27日下午9时和8月17日下午5时、24日下午4时、25日晚上8时、30日下午6时这几个时间段内的传输流量突增。以8月17日的流量异常为例进行分析如图中右侧部分所示，发现10.18网段发出的流量在当日较多，主要是与10.65网段（红色区域）和10.67网段（绿色区域）通信产生的，且向前者传输的流量较多，这与之后的流量突

变相反。更进一步展开右侧流量分析图如图2-5-2所示，发现10.18网段的流量异常主要是由于10.18.112.26向10.65.216.113传输了大量的流量。

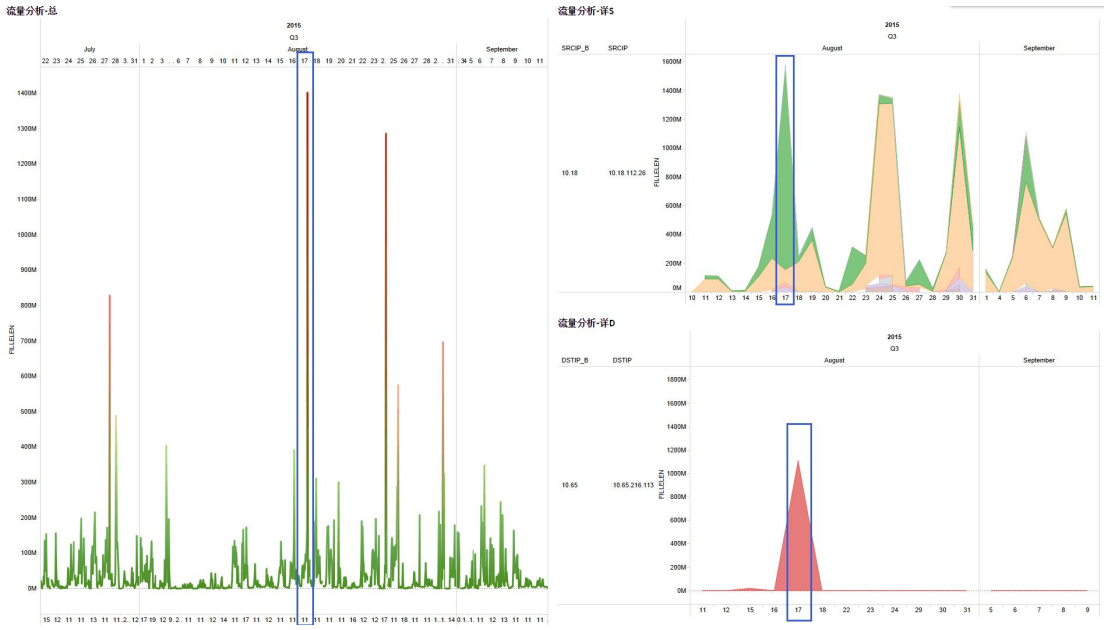

图 2-5-2 10.18 流量异常的 IP

此外，7 月 29 日数据量非常低，几乎没有，9 月 2 日数据缺失。

挑战 1.3： 找出经同一虚拟管道（VPI、VCI）进行通信的源 IP 和目的 IP 的分布规律和连接模式；找出经同一虚拟管道（VPI、VCI）承载的应用分布和变化规律；说明每个（VPI、VCI）虚拟管道的传输数据量在不同时段的变化情况；建议至少给出 5 种虚拟管道通信模式。（请将回答尽量限制在 1500 个字和 10 张图片内）

模式 1：虚拟管道连接的 IP 分布规律及连接模式

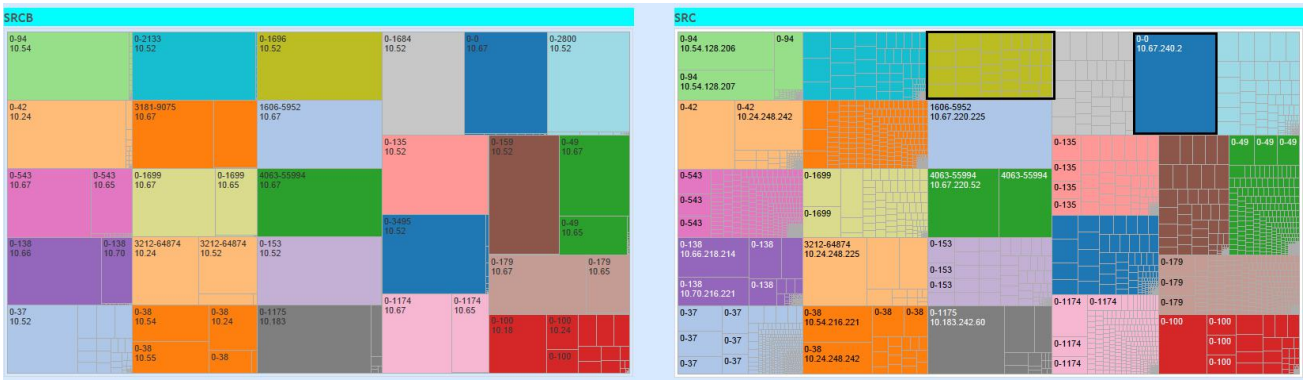

图 3-1-1 虚拟管道的源 IP 分布规律

上图左侧的树地图，不同颜色的各个区域的大小都是相同的，不同的颜色代表不同的虚拟管道，每块颜色区域用使用该管道的源 IP 段分割，分割面积代表该 IP 段使用该管道通信次数的百分比，从左图可以清晰地看到每个管道的源 IP 段分布。右图的树地图和左图的不同在于

每块区域的分割使用的是具体的 IP，从右图可以看出，有些管道的源 IP 分布比较单一，例如使用 0/0 管道（深蓝色）的源 IP 主要是 10.67.240.2，而有的管道的源 IP 分布较复杂，例如 0/1696（黄绿色）管道的源 IP 分布较分散，使用该管道发送信息的 IP 较多且次数较平均。

对虚拟管道的目的 IP 的分布规律分析与对虚拟管道的源 IP 的分布规律分析方法相同，如下图所示：

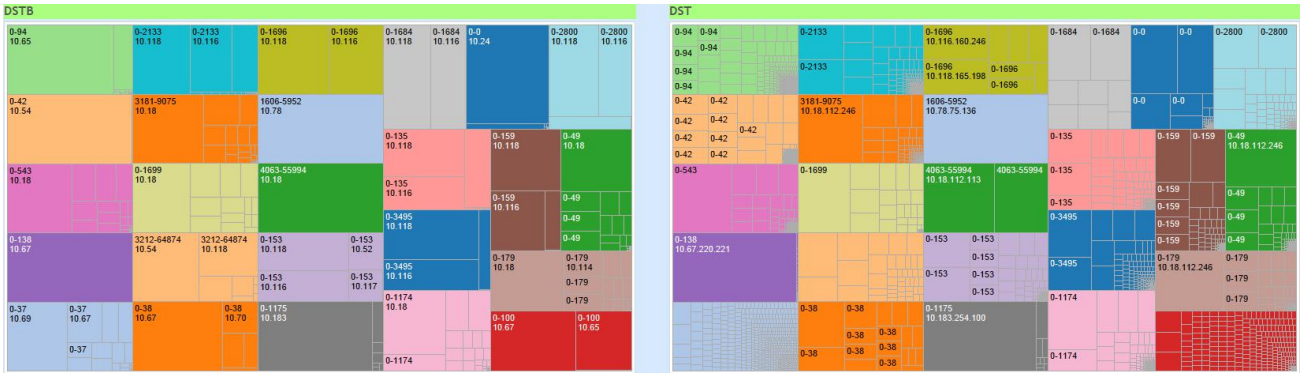

图 3-1-2 虚拟管道的目的 IP 分布规律

对虚拟管道的 IP 连接模式的分析与对虚拟管道的源 IP 分布规律的分析不同的是，对每块区域代表的虚拟管道的分割使用的是 IP 段连接和 IP 连接，如下图所示。从图中可以清晰地看到，每个虚拟管道都主要被一对 IP 段使用，但具体的 IP 对连接分布较分散。例如 0/0 管道主要被 10.67-10.24 使用，但使用该管道的主要 IP 对有 10.67.240.2-10.24.248.242 和 10.67.240.2-10.24.248.253.

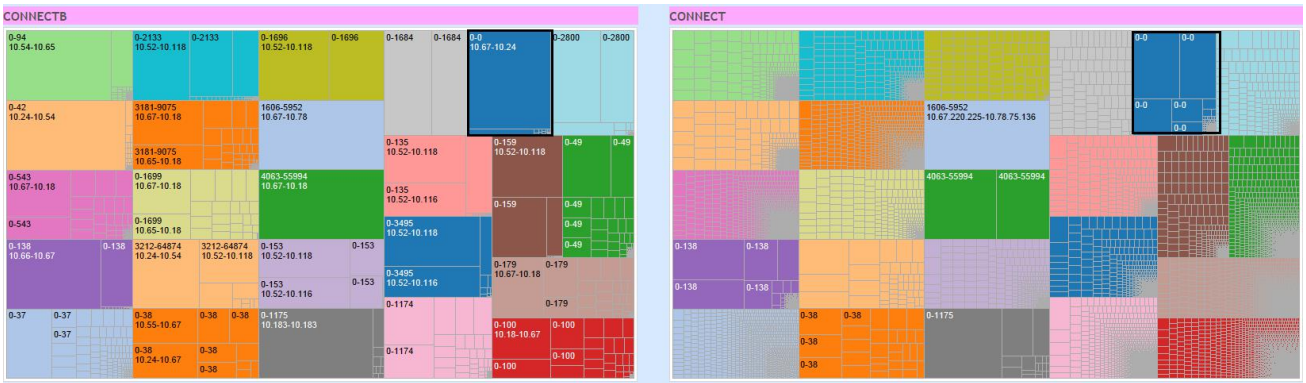

图 3-1-3 虚拟管道的 IP 连接模式

模式 2：虚拟管道承载应用变化模式

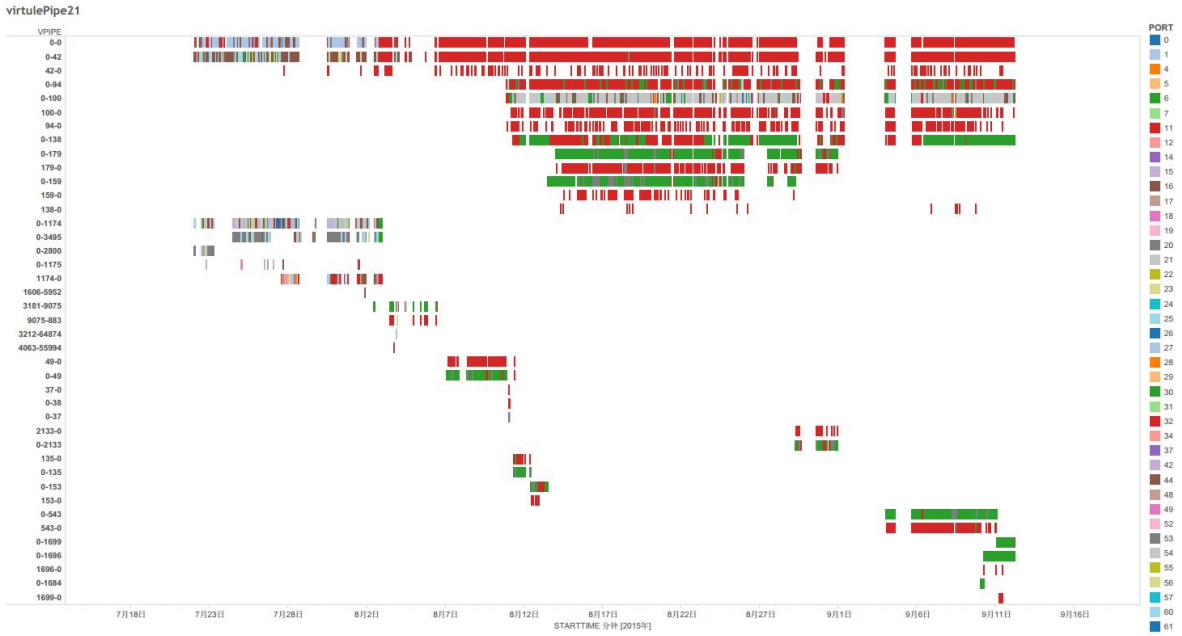

图 3-2-1 虚拟管道承载应用随时间变化

以端口作为应用类型的判定条件，由上图知每个虚拟管道承载的服务随着时间的变化情况。从图中可以看出如下模式：

1）承载最多的应用为端口为 445（红色条带）的应用、端口为 139（绿色条带）的应用，其次为端口为 21（灰色条带）的应用、端口为 135（淡蓝色条带）的应用、端口为 80（棕色条带）的应用和端口为 53（灰色条带）的应用。

2）大部分虚拟管道大部分情况下承载的服务相对稳定，尤其是在 8 月 4 日后，除 0/100 管道承载的服务种类较多，其它管道基本承载相对唯一的服务；8 月 4 日前工作的虚拟管道承载的服务都不唯一，种类较多，推测这是由于企业在 8 月 4 日前业务不稳定，在此之后业务相对稳定，管道承载的应用也相对稳定。

3） 以 0/42 虚拟管道为例分析，8 月 1 日前，该管道大部分情况下主要承载端口为 80 的服务即 http 服务，其它情况下也承载少量的端口号为 25 或 31 的服务；8 月 1 日后，除了 8 月 18 日下午承载了 20 端口的服务外，其它时间只承载 445 端口的服务即文件共享服务。其它虚拟管道承载的服务随时间的变化情况可以结合上图依次得出。

模式 3：虚拟管道流量变化及工作模式

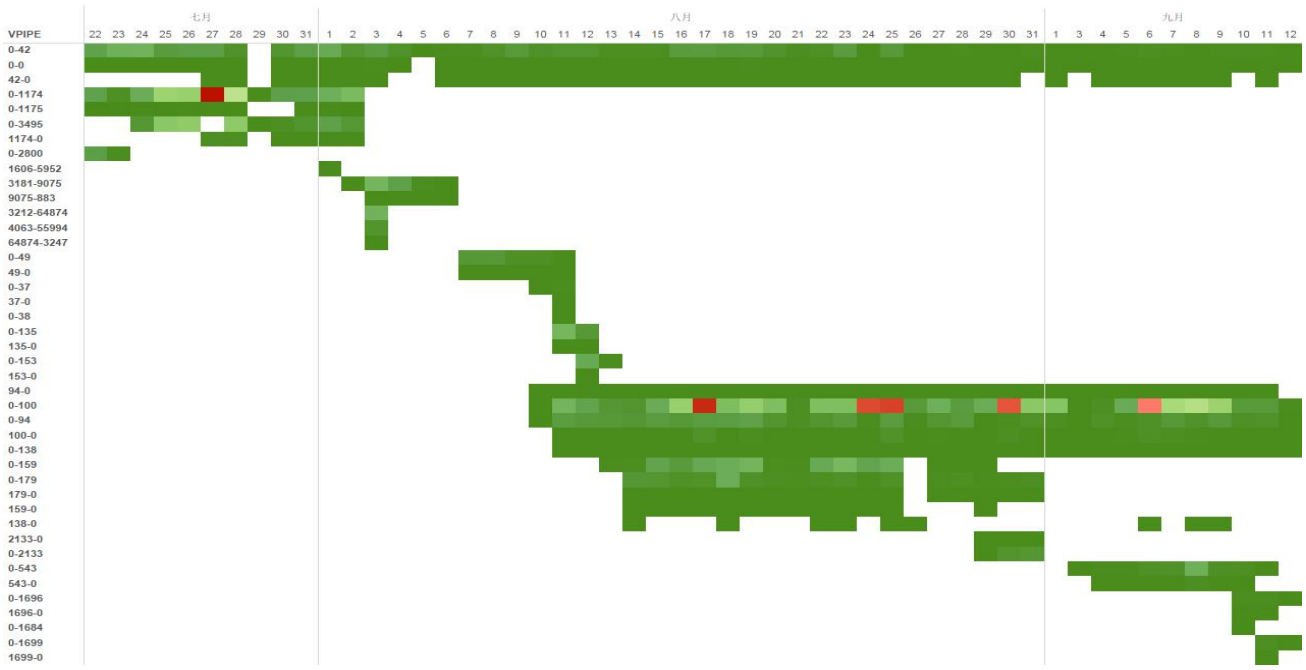

图 3-3-1 虚拟管道流量传输随时间变化情况

对虚拟管道流量传输随时间变化情况1中虚拟管道的顺序进行调整，得到上图。通过上图可以看到，0/1174、0/42和0/3495管道在8月4日前流量变化明显，这主要是因为8月4日前工作的虚拟管道并不多，这三个管道承担了大部分的工作。8月4日之后，0/100、0/94、0/159和0/42管道的流量有明显变化，其中0/100管道在8月17日、25、 26日、30日和9月6日承担的通信次数较平常有很大的增加。其它管道则不然。

此外，由上图可知虚拟管道的工作模式：部分虚拟管道是在整个时间段内都在工作的，如0/42、42/0和0/0管道；有的管道只在8月3日前工作，例如0/1174、0/1175/0/3495和1174/0；有的虚拟管道在8月10日后工作，如94/0、0/94、0/100、100/0、0/138、138/0等管道；有的管道只在9月3日到9月11日工作，例如0/543和543/0管道；除了上述管道，其余管道仅工作几天，推测这些管道可能是为了承担特殊的通信任务而临时配置的。

模式 4：虚拟管道配置模式

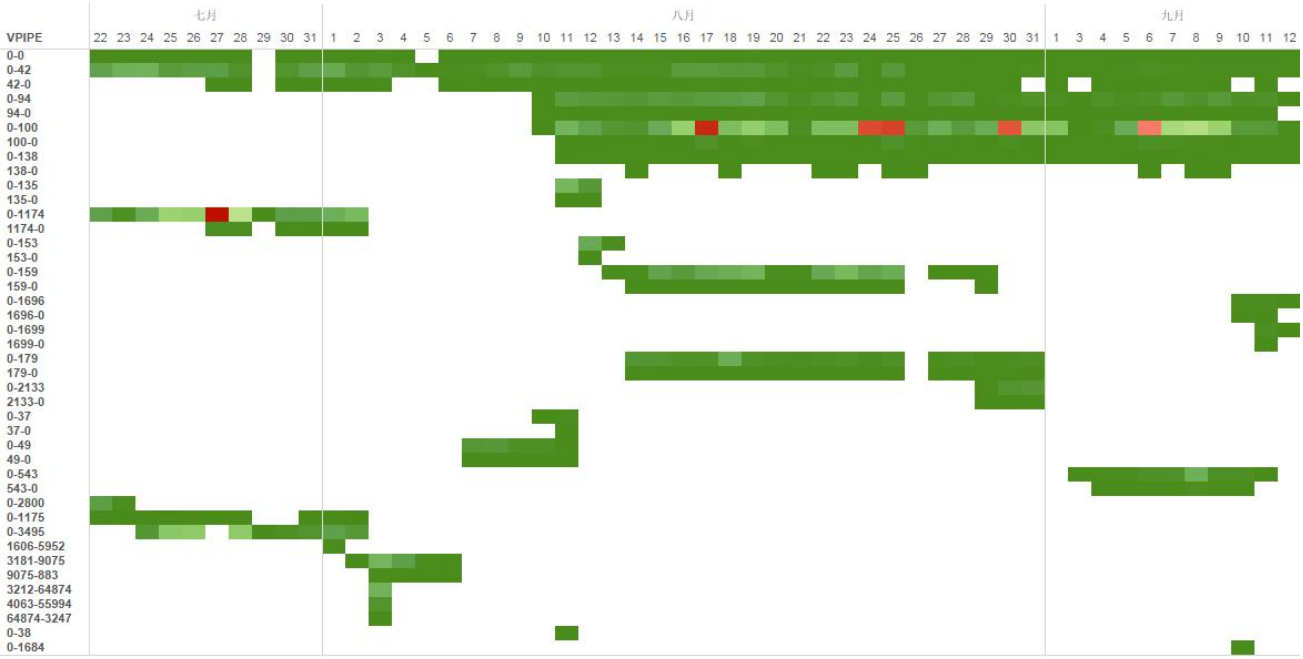

图 3-4-1 虚拟管道流量传输随时间变化情况 1

上图表示虚拟管道传输的流量随着时间的变化情况，由此我们可以推测虚拟管道的配置模式，有的虚拟管道存在与它工作时间相似的管道，这些管道的VPI和VCI是相反的，例如0/94和94/0管道，他们的工作时间都是8月10日以后，0/135和135/0管道，他们的工作时间都是8月11-8月12日这两天，符合这一规律的虚拟管道还有：（0/42, 42/0），(0/100, 100/0)，(0/138, 138/0)等。由此我们推测，为了使用方便，在配置虚拟管道时将工作时间相似的管道的VPI和VCI号设为相反。

模式 5：虚拟管道链路层数据传输模式及文件传输模式

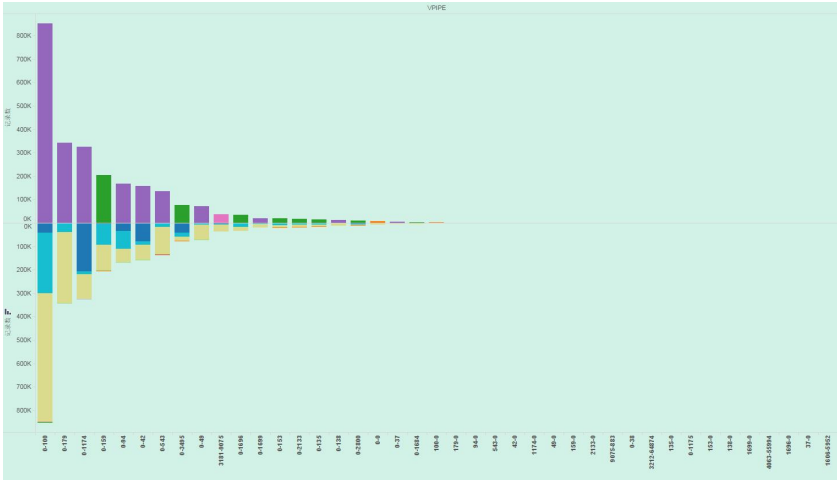

图 3-5-1 虚拟管道的链路层数据类型及传输文件类型

图 3-5-1 中，上方的柱形图表示虚拟管道总的传输流量情况，不同颜色的柱形条表示不同的链路层数据类型。上方各个柱形条颜色唯一，可知各个虚拟管道传输的链路层数据的类型是

唯一的。以 0/42 虚拟管道为例，紫色的条形图代表该管道传输类型为 5 的链路层数据，说明该管道只传输类型为 5 的链路层数据。各虚拟管道传输的链路层数据类型具体如下：

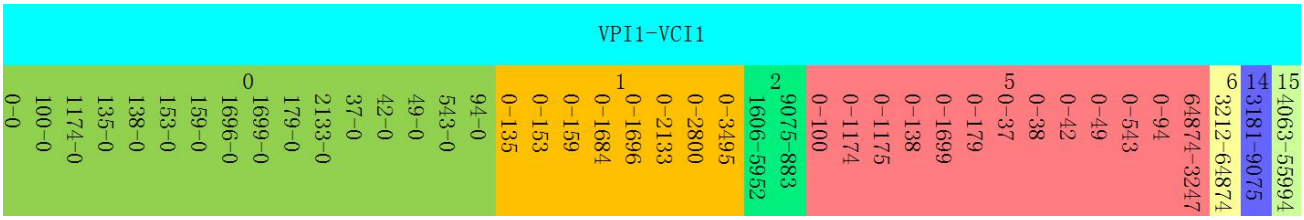

图 3-5-2 虚拟管道传输的链路层数据类型

图3-5-1中，下方的柱形图表示各虚拟管道传输的流量情况，柱形条中的不同颜色部分表示不同的文件类型。由上图各个柱形中不同颜色柱形条的高度，可以看出每个管道主要的文件传输类型。以0/100虚拟管道（下方第一个柱形）为例，该管道主要传输的文件类型为.rpc（蓝色部分）和.null（黄色部分）类型。

模式 6：虚拟管道通信质量变化模式

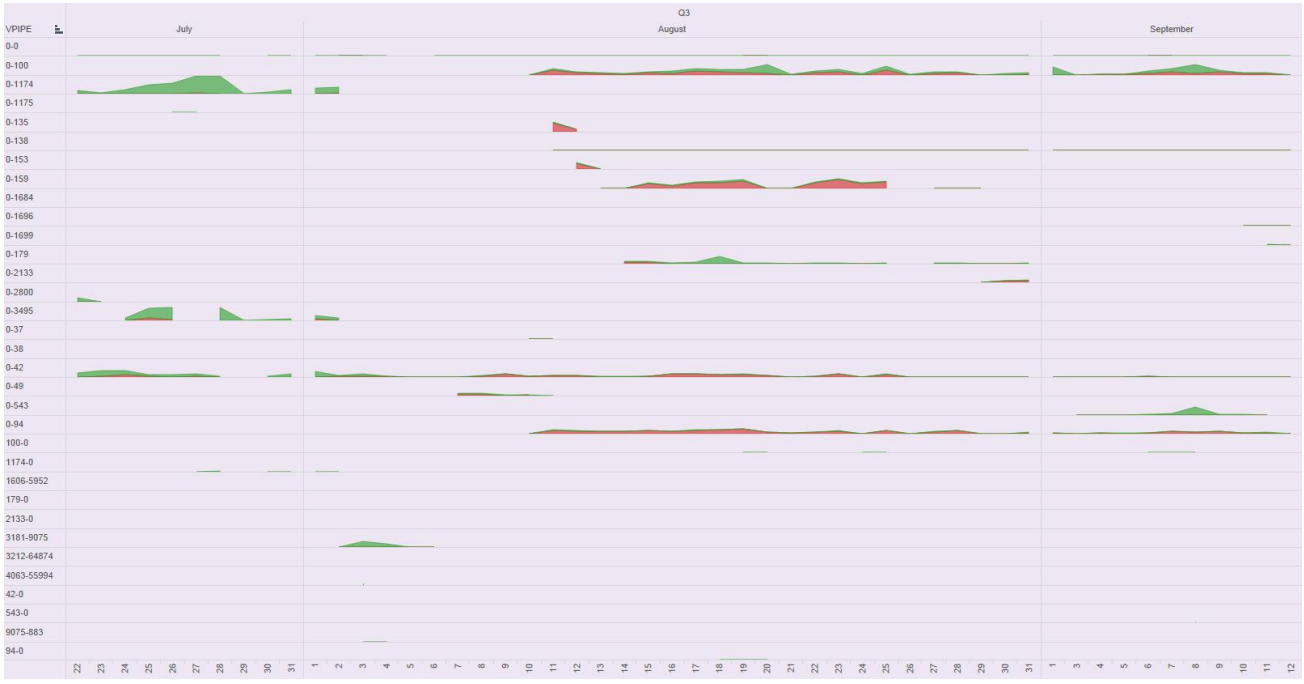

图 3-6-1 虚拟管道通信质量随时间变化情况 2

上图中的每一行代表了一个虚拟管道内的通信次数随时间的变化情况。河流图中的绿色和红色分别表示数据无损、损坏的通信次数。由上图可知每个虚拟管道中传输的数据包被损坏的数量与总量的关系。如从图中可以看出 0/100 管道中通信较频繁的时刻更容易引起较多的数据包被损坏，而 0/1174 的管道单位时间通信次数的增多并未引起被损数据包量的明显增加。总体分析后得出，虚拟管道 0/42、0/100、0/159 和 0/1696 传输数据的失败率随着单位时间内管道被使用频率的增大而迅速增大，说明这些虚拟管道的通信质量不够稳定，其它管道则没有这种情况。

挑战 1.4： WeSuCom 公司要求参赛队设计的可视分析方案能展示数据从源到目的经历应用层、网络层和链路层的全过程，实现对数据连接的多层次可视分析，探索其可能反应的用户行为模式。请参赛队在本题中具体说明其方案是如何满足该需求的。（请将回答尽量限制在 800 个字和 5 张图片内）

对于用户行为的多层次分析，我们的可视方案分为如下四大模块：

1）多用户多层次传输实时模拟

此模块展示某一时间多个用户传输的数据在各个层次的情况。如图4-1-1中所示，右侧的IP表示客户端；中间有管道样式的为虚拟管道；虚拟管道上方为当前时间标注；左边的IP表示服务器；服务器对应的矩形条带表示不同的应用类型；横放的柱形图表示服务器的相应端口截至当前时间的流量累加情况；客户端与虚拟管道、服务器与虚拟管道间连线的颜色表示数据链路层的数据类型。

以8月12日为例进行分析，当日，除了10.75.105.99没有通信，其它客户端都较活跃；用到5个虚拟管道；大多数服务器都比较活跃；链路层数据类型主要有类型5（蓝紫色线条）和类型1（黄绿色线条）；除10.37.216.221和10.18.112.222外，大多数服务器截至当日都有流量的传输。10.52.140.74服务器当日流量传输异常，远大于其它服务器此日传输的流量和此服务器其它时段传输的流量。

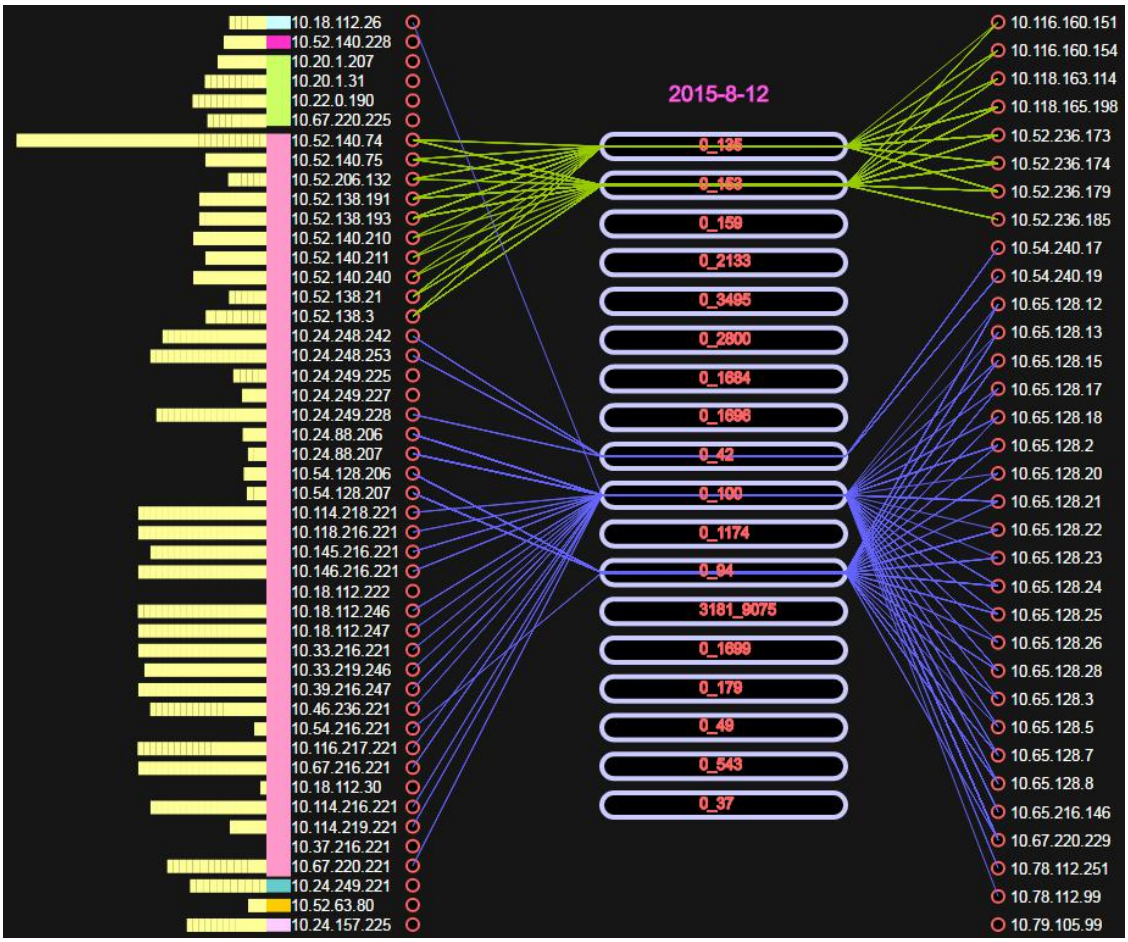

图 4-1-1 多用户多层次传输实时模拟

2）单用户多层次连接实时模拟

此模块能够实时展示每个用户的数据在各个层次间的传输情况。下图中，第一层（以最外层为第一层向里依次递增）是每个用户在整个时间段内传输过的文件类型；第二层为每个用户在整个时间段内传输数据使用的虚拟管道；第三层为每个用户在整个时间段内使用的应用；第四层为用户IP地址；第五层为截至当前时间用户传输的流量累计情况；圆心标注当前时间。由下图可以清晰地看到各个时间每个用户传输的数据在各个层次的情况。

以8月12日为例进行分析，除10.65.216.146外，大部分主机都只使用了一种应用；大部分主机当日使用两个虚拟管道进行数据传输；相比应用和虚拟管道的类型，文件类型较多，各个主机传输的文件类型个数也不一致。以10.52.236.174为例进行分析，该主机当日只使用445文件共享服务，该主机主要通过0/135和0/159管道传输该应用的数据，0/135管道传输的文件类型有.null、.unk、.rpc和.exe，0/159管道传输的文件类型除了上述四种还有.ppt、.xls、.doc和.rtf。

图 4-2-1 单用户多层次传输实时模拟

3）单用户单层次数据传输实时模拟

此模块能够分别实时展示每个IP/流量、IP/通信次数、IP/应用、IP/虚拟管道、IP/文件类型对的数据传输情况，主要使用热度图来实现。以IP/流量和IP/通信次数实时模拟为例进行说明。

IP地址为10.67.220.229的主机在8月15日的流量和通信次数如下图所示，分别为54.2M和3815次，下图的左半部分从上至下从左至右依次为该主机当天各个小时的通信次数情况，与服务器之间通信的流量情况，与其它主机通信的流量情况。玫瑰图按照角度将当天的24小时平均分为4个时段，每个时段6个小时，每个时段又被分成内外两部分，内层描黄边的环的半径正比于此小时内该主机被连接的次数，外层描红边环的半径正比于该主机发起通信的次数。左下角的弧形图表示该主机作为目的主机时的连接情况，右下角弧形图表示该主机向其它主机发送的流量情况弧的宽度表示流量的大小。

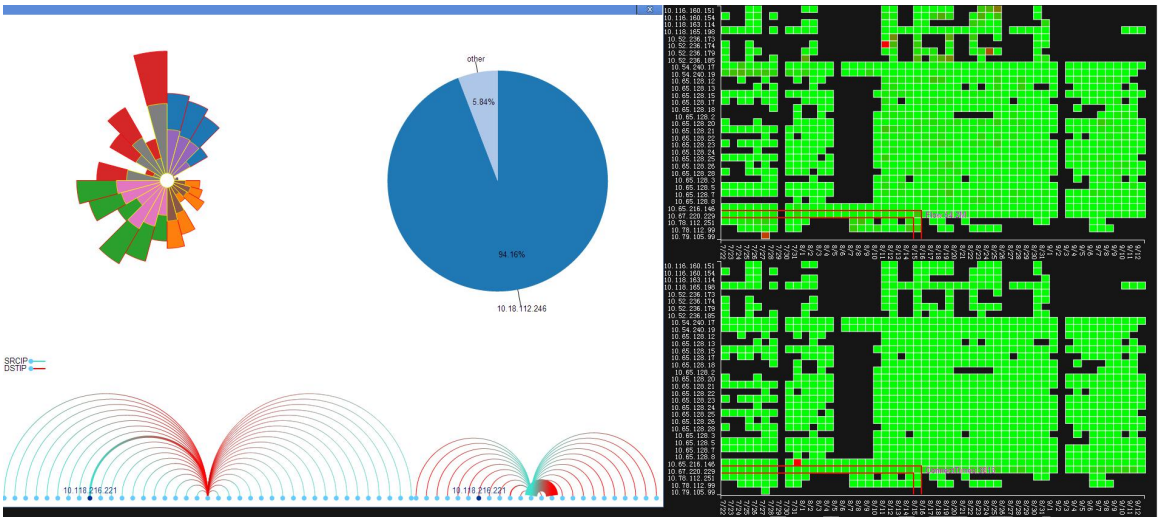

图 4-3-1 10.67.220.229 在 8 月 15 日的 IP/流量和 IP/通信次数情况

4）多层次间属性关系总体分析

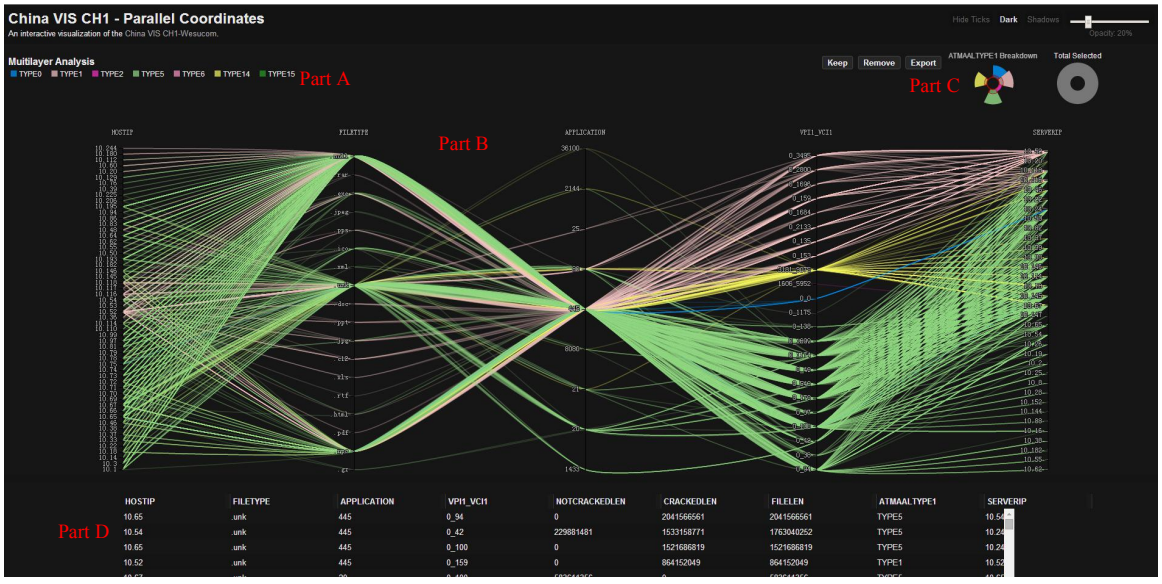

图 4-4-1 多层次多属性关系平行坐标系

此模块能够展示多个层次的不同属性之间的联系，主要通过结合了玫瑰图的平行坐标系来实现。如图 4-4-1 所示，Part A 标识平行坐标系中线的颜色与数据链路层 ATMAAL1TYPE 属性的七种不同取值的对应关系。Part B 中的平行坐标系展示不同网段的主机与服务器通信时所使用的文件类型、端口号与虚拟管道。Part C 的玫瑰图展示 ATMAAL1TYPE 值为 0、1、2、5、

6、14、15 的各种数据流量的大小，以内层描红色边的环和外层不描边的环分别表示传输失败与成功的流量所占的比重；Part C 的单层环图展示符合 Part B 中被选中条件的数据占总记录的百分比。Part D 显示数据的具体情况。用户可以用刷子在平行坐标系的一个或多个轴上选取任意的范围进行分析。Part D 中被选中的数据会在平行坐标系中高亮。

从图4-4-1中发现1）主机与服务器通信大多是通过445端口发送或接收文件类型为:.unk,.rpc以及.unll(空类型)的数据，这些数据大多是通过VPI1=0,VCI=*的各虚拟管道进行传输的；2）由玫瑰图可知：ATMAAL1TYPE值为1、5、14类型数据的流量明显比较多且大部分都能成功传输，而ATMAAL1TYPE值为0、2、6、15类型数据的流量明显很少，但它们几乎全部传输成功；3）在VPI1_VCI1轴上黄色的线只分布于3181_9075的刻度上，这说明TYPE14只使用VPI1=3181 VPI1=9075的虚拟管进行传输。ATMAAL1TYPE其它类型的数据使用多种虚拟管道进行传输。4）一个虚拟管道只传输一种ATMAAL1TYPE类型的数据（VPI1_VCI1轴上的每一个值只对应一种颜色的线），且值为0的数据大多通过VPI1=0,VCI1=0的虚拟管道传输。

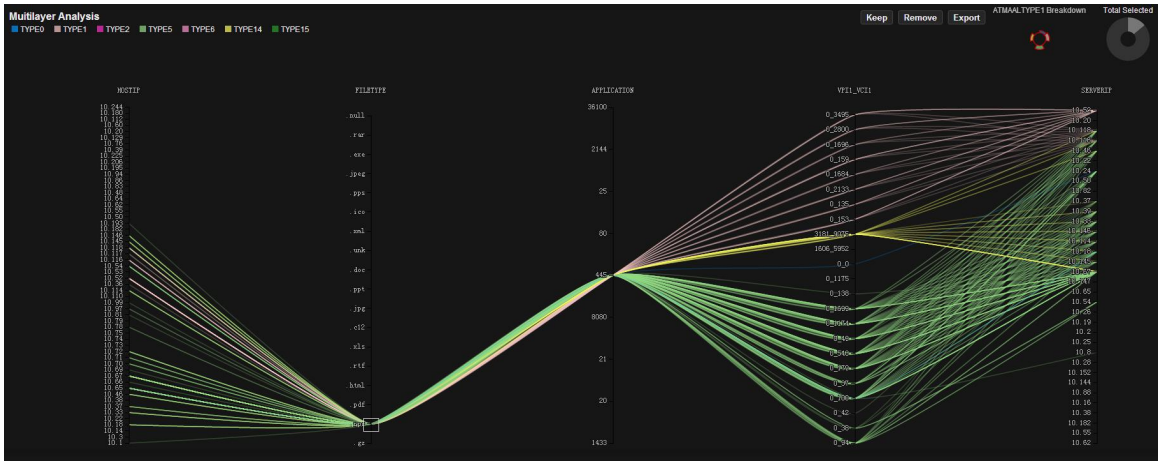

图 4-4-2 .rpc 类型文件与其它层次属性的关系

通过平行坐标系的交互操作可以分析出各个属性间的关联。如图 4-4-2 所示选中 FILETYPE 轴上的.rpc，发现这种文件类型主要通过端口为 445 的应用传输，传输该类文件的主机与服务器分布在多个网段，使用的虚拟管道也较多，但根据玫瑰图发现，该文件全部传输失败，由单层环图可知这样的记录并不多。