2018 年第五届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 2018 - mini challenge 1）

答　卷

参赛队名称：成都理工大学-冯晓蕾-挑战 1

团队成员：　冯晓蕾，成都理工大学，812934728@qq.com，队长

成利，成都理工大学，841482039@qq.com

秦阳欣，成都理工大学，531253432@qq.com

李浙川，成都理工大学，827778900@qq.com

林杰，成都理工大学，598867407@qq.com，指导老师

团队成员是否与报名表一致（是或否）：是

是否学生队（是或否）：　是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）：Excel，Pyecharts, Tableau,Gephi

共计耗费时间（人天）：　40 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：分析公司内部员工所属部门及各部门的人员组织结构，给出公司员工的组织结构图。

图 1-1- 1 部门关键词

图 1-1-1 表示每个部门发送邮件的主题，每个部门发送邮件主题都不尽相同，根据邮件内容的不同，可以区分员工属于哪个部门。例如：部署文档，地图配置，软件开发文档等，都是研发部门发送邮件的关键词，可以判断发送该邮件内容的是研发部门人员。最后结果显示人力资源部门 19 人，研发部门 256 人，财务部门 24 人，共计 299 人。

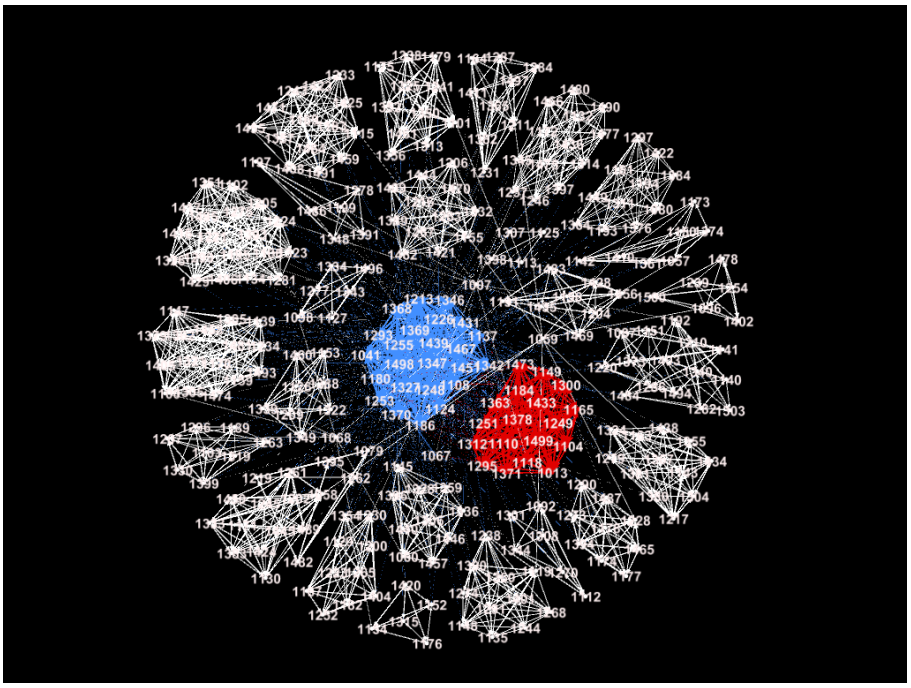


图 1-1-2 公司内部邮件收发关系图

图 1-1-2 表示公司内部的邮件收发状况，其邮件的收发情况即形成了一个网络，利用社区发现算法进行分析，蓝色表示为财务部门，红色表示为人力资源部门，白色为研发部门，其中研发部门又分为若干个小组。

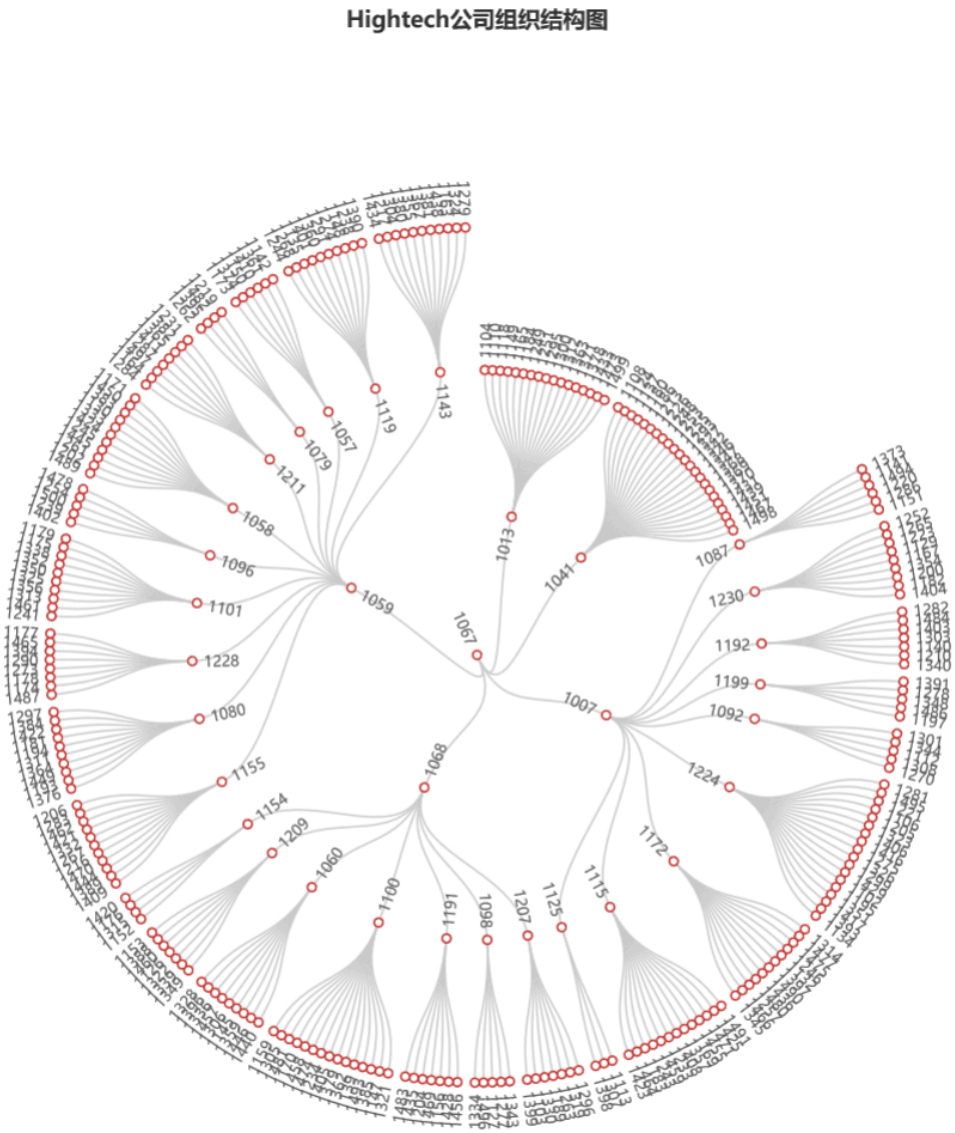


图 1-1-3 员工组织结构图

图 1-1-3 表示公司员工组织结构图，中间为公式老板，往外依次是部门负责人、研发部项目组组长、员工。财务部负责人为 1041，人力资源部负责人为 1013，研发部门负责人为 1007、1059、1068。

项目组 1007 组长：1192、1172、1125、1092、1199、1087、1115、1224、1230

项目组 1059 组长：1119、1155、1079、1057、1143、1101、1211、1096、1080、1058、1228

项目组 1068 组长：1060、1191、1207、1154、1209、1100、1098

挑战 1.2：分析该公司员工的日常工作行为，按部门总结并展示员工的正常工作模式。

（一）工作时间

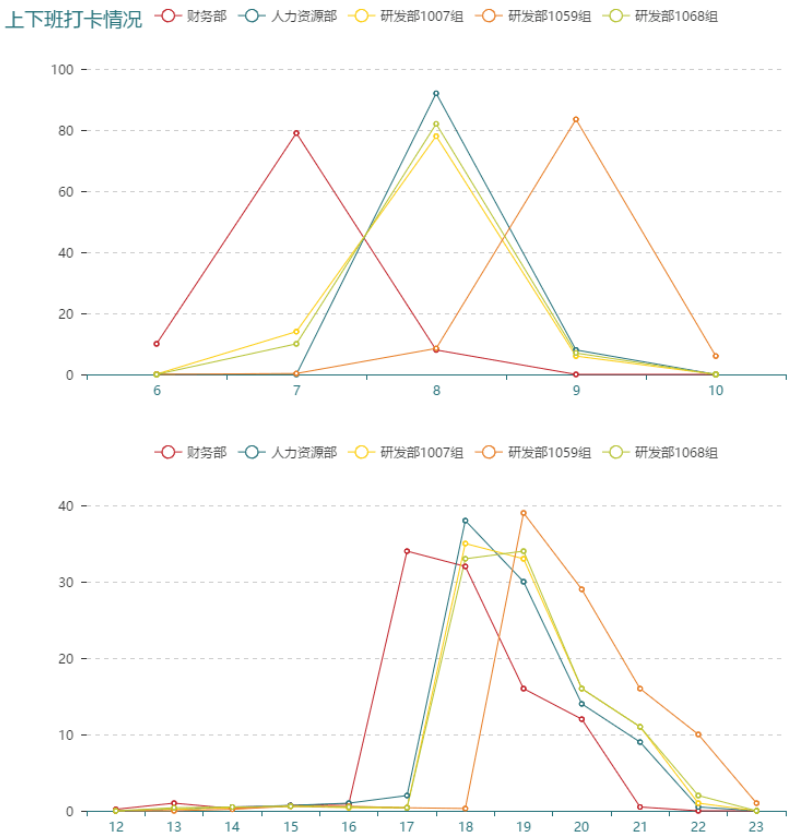


图 1-2-1 部门签到及签离

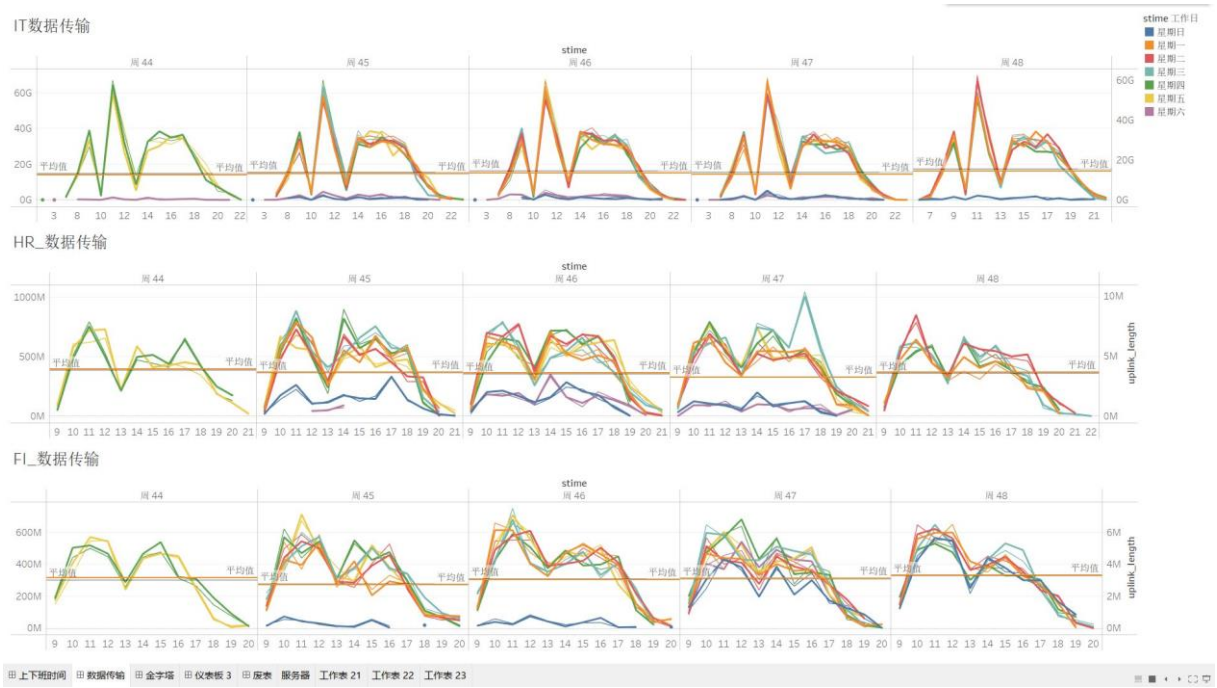


图 1-2-2 数据流量传输图

图 1-2-1 表示各部门的签到及签离情况，折线图图的高度表示员工签到或签离的时间段的人数，折线图的颜色表示部门，由图可以知道各部门的上下班时间。

图 1-2-2 表示各个部门的流量传输情况，可以看到每天 12 点都存在一个较大的波峰，紧接着会出现波谷，可以推测为其公司的中午休息时间。

（1）研发部门工作时间：

项目组 1007 工作时间：09：00~12：0014：00~18：00

项目组 1059 工作时间：10：00~12：0014：00~19：00

项目组 1068 工作时间：09：00~12：0014：00~18：00

（2）人力资源部门工作时间：09：00~12：0014：00~18：00

（3）财务部门工作时间：08：00~12：0014：00~17：00

（二）工作内容

图 1-2-3 部门邮件关键词

如图 1-2-3 所示，研发部工作内容：研发部主要负责软件的开发，包括需求分析、需求调研、项目计划、问题、风险、软件实现、测试、技术分析等，其内容无具体分工。

财务部工作内容：财务部的工作内容包括财务分析、成本控制、资金、会计核算、税务、财务报账等。

人力资源部工作内容：人力资源的工作内容包括招聘、效绩考核、发送劳动合同、考勤、福利保障等。

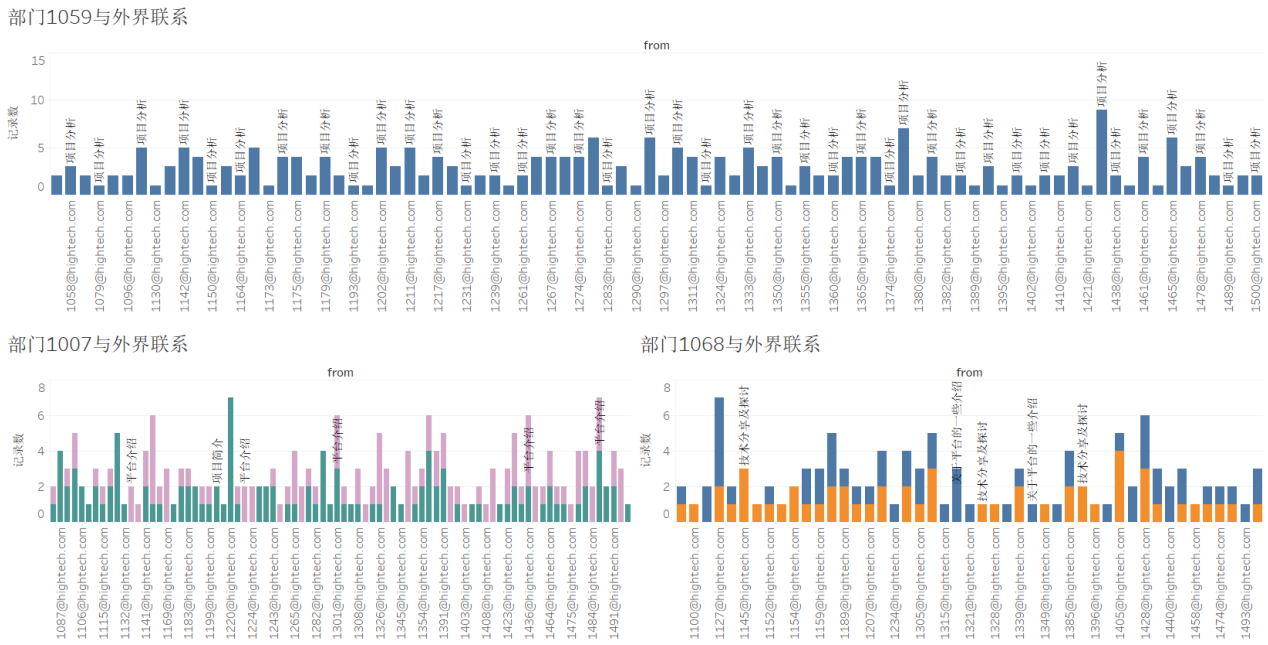


图 1-2-4 研发部每个项目组与外界交流

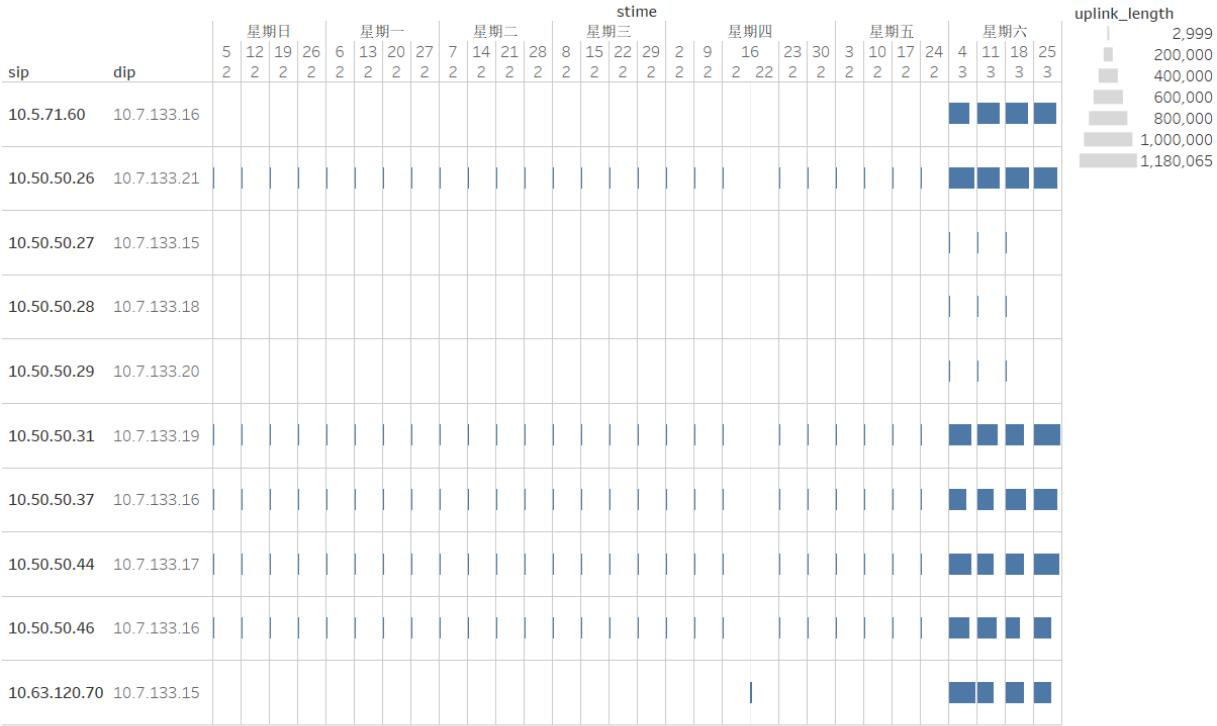
如图 1-2-4 所示，项目组 1007 负责项目分析，项目组 1059 负责项目简介和平台简介，项目组 1068 负责平台介绍和技术分析及探讨。

（三）服务器更新

Server								
tds	ssh	smtp	sftp	postgresql	mysql	mongodb	http	ftp
10.50.50.49	10.63.120.70	10.7.133.21	10.50.50.49	10.50.50.49	10.50.50.49	10.50.50.49	10.63.120.70	10.50.50.49
10.50.50.48	10.50.50.49	10.7.133.20	10.50.50.48	10.50.50.48	10.50.50.48	10.50.50.48	10.50.50.49	10.50.50.48
10.50.50.47	10.50.50.48	10.7.133.19	10.50.50.47	10.50.50.47	10.50.50.47	10.50.50.47	10.50.50.48	10.50.50.47
10.50.50.46	10.50.50.47	10.7.133.18	10.50.50.46	10.50.50.46	10.50.50.46	10.50.50.46	10.50.50.47	10.50.50.46
10.50.50.45	10.50.50.46	10.7.133.17	10.50.50.45	10.50.50.45	10.50.50.45	10.50.50.45	10.50.50.46	10.50.50.45
10.50.50.44	10.50.50.45	10.7.133.16	10.50.50.44	10.50.50.44	10.50.50.44	10.50.50.44	10.50.50.45	10.50.50.44
10.50.50.43	10.50.50.44	10.7.133.15	10.50.50.43	10.50.50.43	10.50.50.43	10.50.50.43	10.50.50.44	10.50.50.43
10.50.50.42	10.50.50.43	10.5.71.60	10.50.50.42	10.50.50.42	10.50.50.42	10.50.50.42	10.50.50.43	10.50.50.42
10.50.50.41	10.50.50.42		10.50.50.41	10.50.50.41	10.50.50.41	10.50.50.41	10.50.50.42	10.50.50.41
10.50.50.40	10.50.50.41		10.50.50.40	10.50.50.40	10.50.50.40	10.50.50.40	10.50.50.41	10.50.50.40
10.50.50.39	10.50.50.40		10.50.50.39	10.50.50.39	10.50.50.39	10.50.50.39	10.50.50.40	10.50.50.39
10.50.50.38	10.50.50.39		10.50.50.38	10.50.50.38	10.50.50.38	10.50.50.38	10.50.50.39	10.50.50.38
10.50.50.37	10.50.50.38		10.50.50.37	10.50.50.37	10.50.50.37	10.50.50.37	10.50.50.38	10.50.50.37
10.50.50.36	10.50.50.37		10.50.50.36	10.50.50.36	10.50.50.36	10.50.50.36	10.50.50.37	10.50.50.36
10.50.50.35	10.50.50.36		10.50.50.35	10.50.50.35	10.50.50.35	10.50.50.35	10.50.50.36	10.50.50.35
10.50.50.34	10.50.50.35		10.50.50.34	10.50.50.34	10.50.50.34	10.50.50.34	10.50.50.35	10.50.50.34
10.50.50.33	10.50.50.34		10.50.50.33	10.50.50.33	10.50.50.33	10.50.50.33	10.50.50.34	10.50.50.33
10.50.50.31	10.50.50.33		10.50.50.31	10.50.50.31	10.50.50.31	10.50.50.31	10.50.50.33	10.50.50.31
10.50.50.30	10.50.50.31		10.50.50.30	10.50.50.30	10.50.50.30	10.50.50.30	10.50.50.31	10.50.50.30
10.50.50.29	10.50.50.30		10.50.50.29	10.50.50.29	10.50.50.29	10.50.50.29	10.50.50.30	10.50.50.29
10.50.50.28	10.50.50.29		10.50.50.28	10.50.50.28	10.50.50.28	10.50.50.28	10.50.50.29	10.50.50.28
10.50.50.26	10.50.50.28		10.50.50.26	10.50.50.26	10.50.50.26	10.50.50.26	10.50.50.28	10.50.50.26
10.7.133.22	10.50.50.26		10.7.133.22	10.7.133.22	10.7.133.22	10.7.133.22	10.50.50.27	10.7.133.22
10.7.133.21	10.7.133.22		10.7.133.21	10.7.133.21	10.7.133.21	10.7.133.21	10.50.50.26	10.7.133.21
10.7.133.20	10.7.133.21		10.7.133.20	10.7.133.20	10.7.133.20	10.7.133.20	10.7.133.22	10.7.133.20
10.7.133.19	10.7.133.20		10.7.133.19	10.7.133.19	10.7.133.19	10.7.133.19	10.7.133.20	10.7.133.19
10.7.133.18	10.7.133.19		10.7.133.18	10.7.133.18	10.7.133.18	10.7.133.18	10.7.133.19	10.7.133.18
10.7.133.16	10.7.133.18		10.7.133.16	10.7.133.16	10.7.133.16	10.7.133.16	10.7.133.18	10.7.133.16
10.7.133.15	10.7.133.17		10.7.133.15	10.7.133.15	10.7.133.16	10.7.133.15	10.7.133.16	10.7.133.15
	10.7.133.16						10.7.133.15	
	10.7.133.15						10.5.71.60	
	10.5.71.60							

图 1-2-5 服务器

服务器数据更新



按 stime 工作日, stime 天 与 stime 小时 以及 sip 与 dip 细分的 uplink_length 总计 (大小)。在 sip 与 dip 上筛选视图。sip 筛选器保留 10 个成员。dip 筛选器保留 7 个成员(总共 366 个)。

图 1-2-6 服务器数据更新

研发部服务器：

10.50.50.26（git） 10.50.50.27（jira） 10.50.50.28（lib01） 10.50.50.29（lib02）（主要在用）10.50.50.31 10.50.50.37 10.50.50.44 10.50.50.46，其服务器会在每周六凌晨 03：00 将数据上传，除周六每天晚上 02:00 会将少量数据上传。

财务部服务器：

10.63.120.70（OA）（主要在用）10.5.71.60（email），均为公司服务器，其服务器会在每周六凌晨 03：00 将数据上传，除周六每天晚上 02:00 会将少量数据上传。

人力资源部服务器：

10.63.120.70 （OA）10.5.71.60（email）（主要在用），均为公司服务器，其服务器会在每周六凌晨 03：00 将数据上传，除周六每天晚上 02:00 会将少量数据上传。

（四）工作模式

图 1-2-7 结果

最后结果采用树图的方式输出，点击相应的部门即可查看各个部门的工作总结、辞职流程以及入职流程等工作情况。

挑战 1. 3：找出至少 5 个异常事件，并分析这些事件之间可能存在的关联，总结你认为有价值的威胁情报，并简要说明你是如何利用可视分析方法找到这些威胁情报的。

（一）登录失败异常

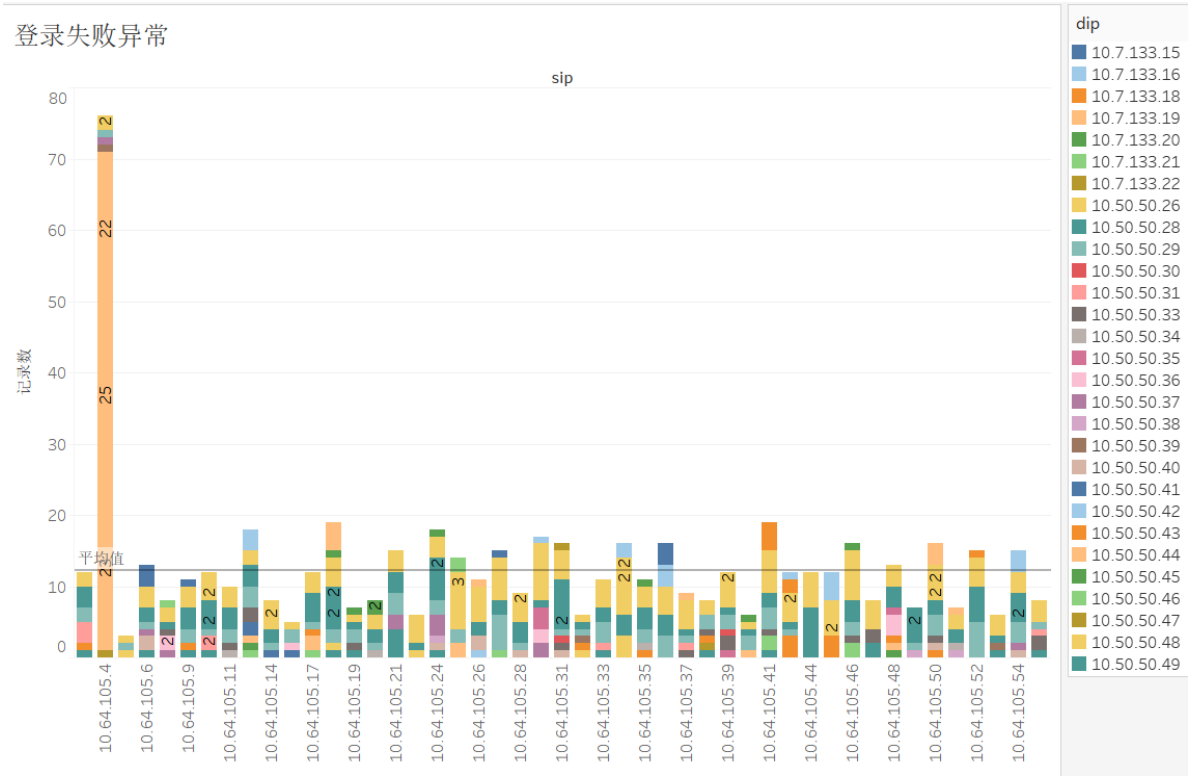


图 1-3-1 登录失败异常

如图 1-3-1 所示，使用 Tableau 制作条形图，横坐标为源 IP，纵坐标为累计登录失败次数，颜色表示目的 IP，分段数字表示某天次数，条形图的高度表示次数的多少。IP 为 10.64.105.4 的用户 1487 登录失败异常， 其中 3 号 22 次，4 号 25 次，6 号 23 次，可以推断其行为异常，可能在做超出其权限的登录。

（二）服务器更新异常

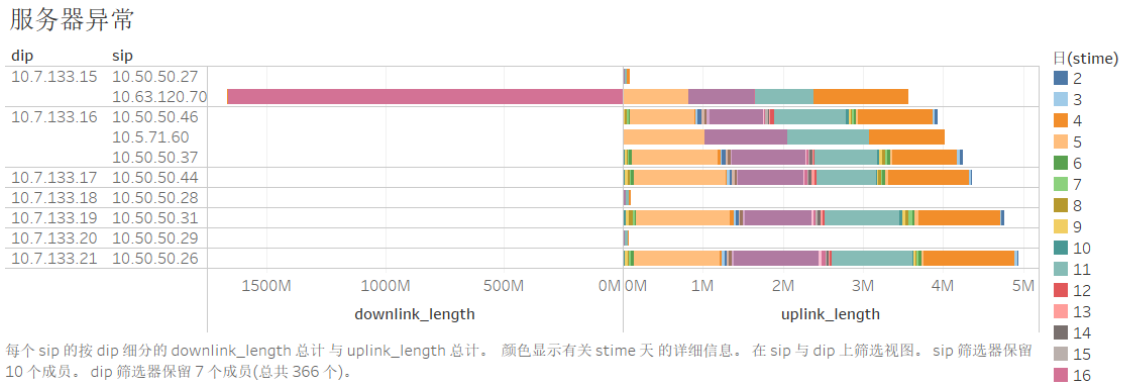


图 1-3-2 服务器异常

如图 1-3-2 所示，利用 Tableau 制作金字塔图，纵坐标表示源 IP 和目的 IP，横坐标是由上传量和下载量制作的双轴，颜色表示时间，条形图的长短表示多少。IP 为 10.63.120.70 的服务器在 11 月 16 日存在一个异常下载，而当日 alert@hightech.com 给 1487 和 1284 发送“EmergencyDataBaseFatalError!”。

（三）垃圾邮件

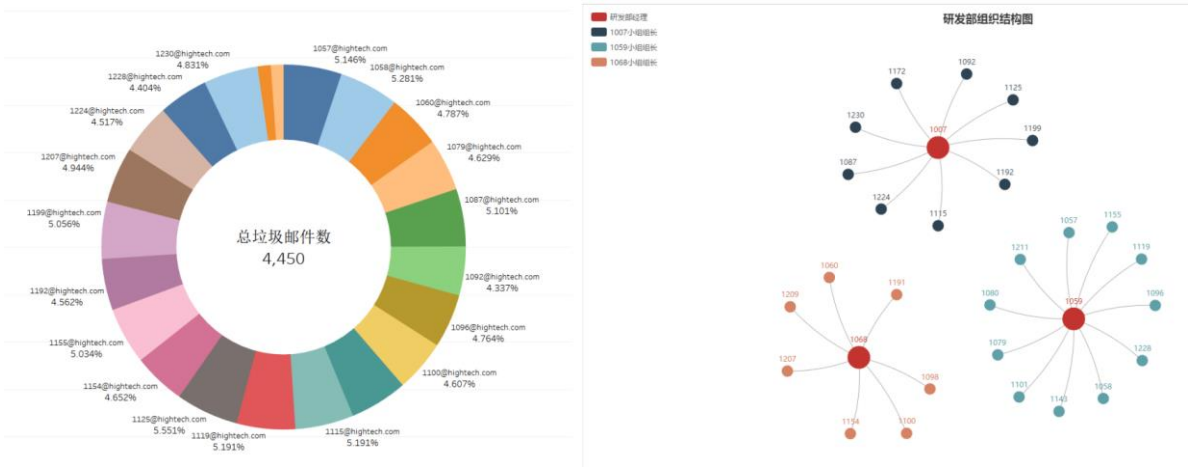


图 1-3-3 垃圾邮件

如图 1-3-3 所示，左边圆环图表示员工接受垃圾邮件所占百分比，右边关系图使用 Echarts 制作，节点表示员工 ID，大小表示职位大小，线表示存在上下属关系。垃圾邮件都是固定发给一些人的，而这些人全部为研发部的小组长，而垃圾邮件会导致安全邮件崩溃。

（四）安全邮件崩溃异常

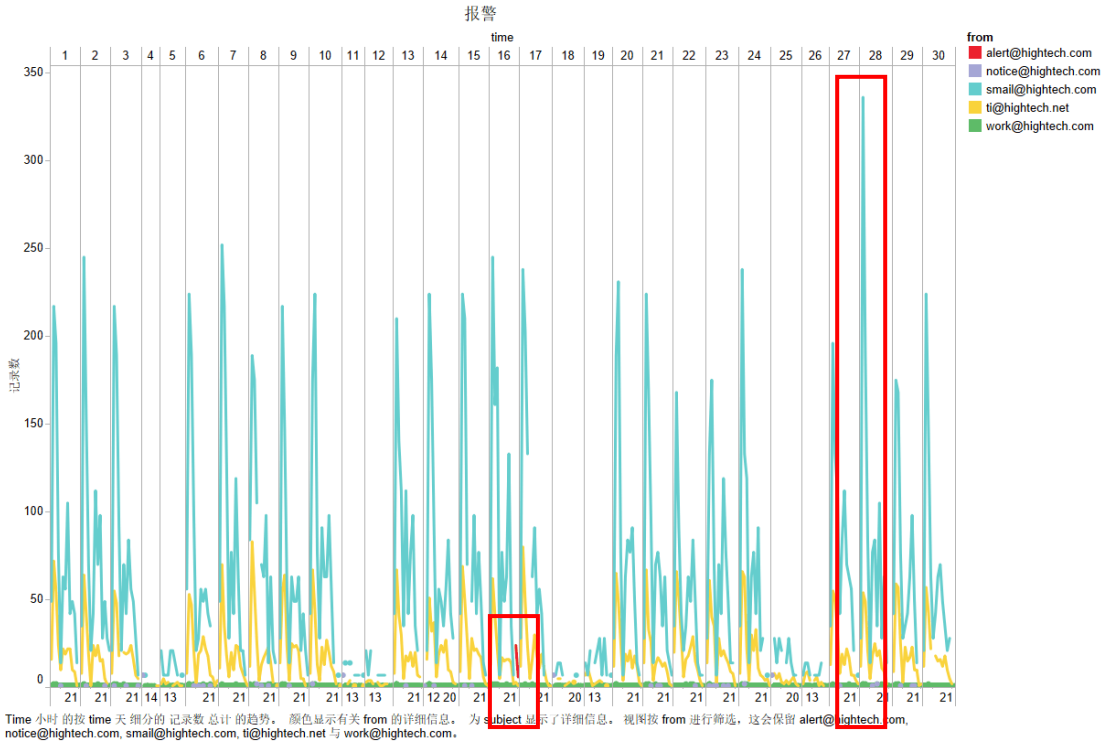


图 1-3-4 报警

如图 1-3-4 所示，折线图的高低表示收到报警的次数，横轴表示时间天和小时。邮箱内容存在五种警报，ti@hightech.net 发送“互联网资产监控报警”，smail@hightech.com 发送“安全邮件崩溃”，work@hightech.com 发送“[ALARM:100]HOST...”，notice@hightech.com 发送通知、提醒，alert@hightech.com 发送“EmergencyDataBaseFatalError!”，28 号那天的安全邮件崩溃比平时高很多。

（五）数据库崩溃异常

如图 1-3-4 所示，数据库崩溃只发生在 16 号，且数据库第二天可以正常使用。

（六）招聘网页浏览

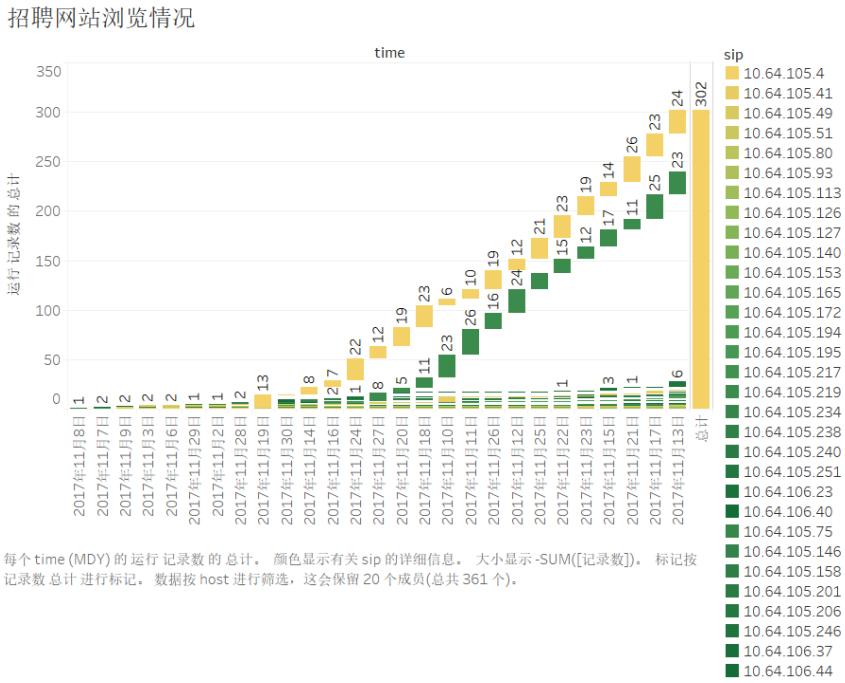


图 1-3-5 招聘网站浏览

如图 1-3-5 所示，瀑布图的高度表示为浏览招聘网页次数的多少，颜色表示源 IP，其中 IP 为 10.64.105.4 的员工 1487 和 IP 为 10.64.105.219 的员工 1376 访问次数异常。

（七）服务器活动异常

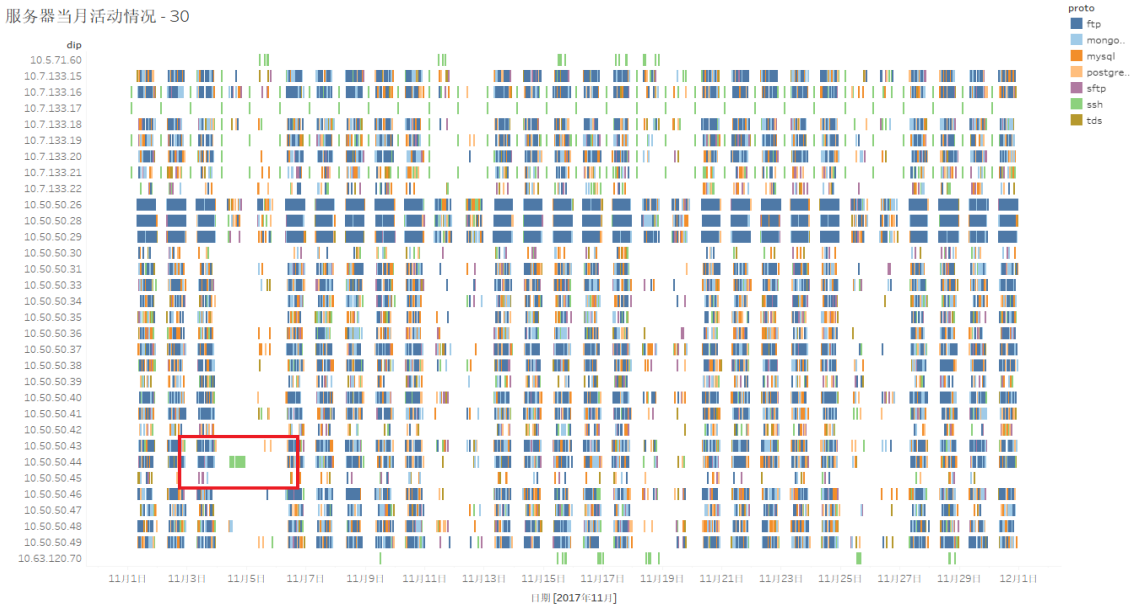


图 1-3-6 服务器活动异常

如图 1-3-6 所示，条形图长短表示次数，颜色表示使用协议，纵轴为目的 IP，横轴为时间。在非工作日 11 月 4 日，IP 地址为 10.50.50.44 的服务器存在大量活动异常情况，访问人 IP 为 10.64.105.4，访问人为 1487，且传输协议均为 ssh。

（八）员工访问异常

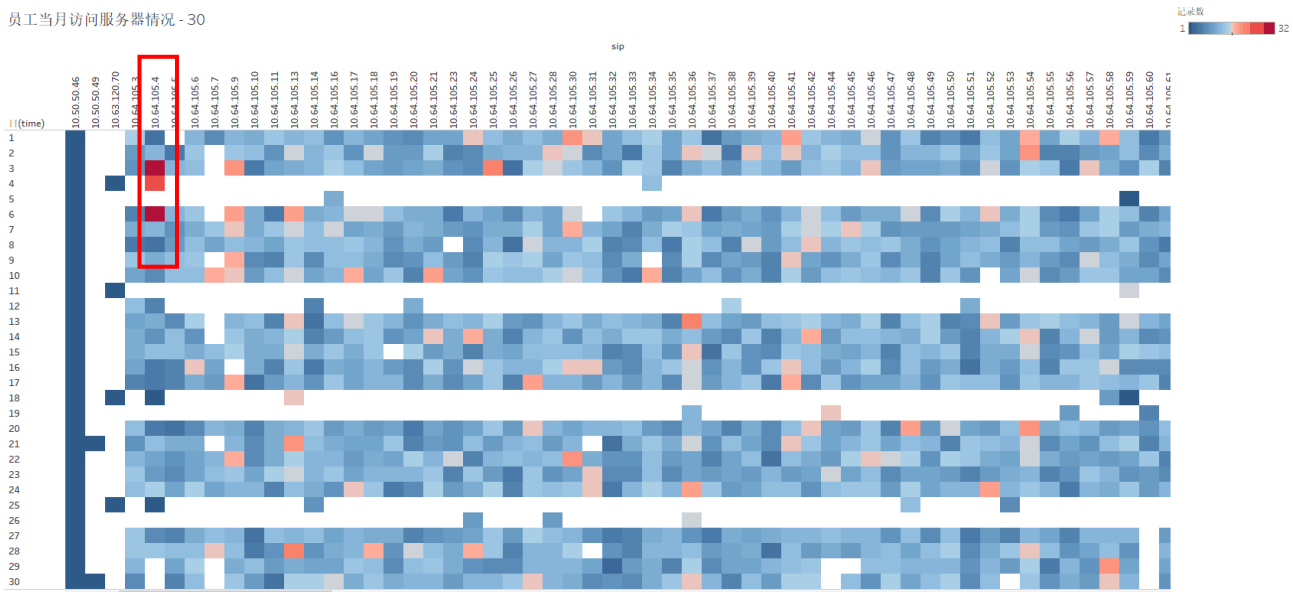


图 1-3-7 员工访问异常

如图 1-3-7 所示，横轴为目的 IP，纵轴为时间，颜色表示次数。IP 地址为 10.64.105.4 的员工在 3、4、6 号这三天对服务器的访问量出现大量增加的异常情况。

联系：

(1)登录失败异常的用户为 1487，而 1487 也是接受数据库崩溃的人员之一，管理服务器的人之一，而服务器崩溃后有一次异常的数据更新。

(2)垃圾邮件的接收会导致安全邮件崩溃，而接收安全邮件的均是研发部门的小组长。

(3)异常（一）、（六）、（七）明显为员工 1487 对 10.50.50.44 服务器的大量异常访问活动造成。

（九）因此，根据找到的异常情况，我们可以向该公司提出以下几个方案保证信息安全：

首先，是保护员工电脑文件安全、防止公司数据泄密。

由于员工的日常工作都是通过计算机网络实现，工作中形成的重要文件也常常存储在员工电脑上，这样就使得员工可以轻松通过 U 盘、移动硬盘等 USB 存储设备复制电脑文件，或者通过邮件、网盘、FTP 文件上传或者 QQ 发送文件、微信发送文件的方式将电脑文件发送出去，从而盗取了公司商业机密信息。为此，我们需要对企业局域网重要的员工电脑，禁用 USB 存储设备，并禁止上述网络泄密通道，同时，在员工离职之前，要检查并保证员工没有携带公司的情报离职。其次，还需要防止局域网共享文件服务器泄密的行为。

由于很多单位局域网都有文件服务器，通常会将单位重要的文件共享给局域网访问使用，便于实现协同办公。但由于缺乏对用户访问权限的管控，使得共享文件泄密也极为方便。因为，只要允许用户读取共享文件就意味着用户可以轻松复制共享文件内容，并且还可以将共享文件另存为本地磁盘，或者拖拽共享文件到访问者自己的电脑。此外，用户如果具有共享文件修改权限的情况下，还可能不小心或故意删除共享文件，从而给共享文件安全带来巨大隐患。

最后，要防止垃圾邮件攻击。

所谓的垃圾邮件一般具有批量发送的特征。其内容包括赚钱信息、成人广告、商业或个人网站广告、电子杂志、连环信等。垃圾邮件可以分为良性和恶性的。良性垃圾邮件是各种宣传广告等对收件人影响不大的信息邮件。恶性垃圾邮件是指具有破坏性的电子邮件。

有些垃圾邮件发送组织或是非法信息传播者，为了大面积散布信息，常采用多台机器同时巨量发送的方式攻击邮件服务器，造成邮件服务器大量带宽损失，并严重干扰邮件服务器进行正常的邮件递送工作无论垃圾邮件的下一步会怎么走，企业都应该部署好可以侦测并封锁电子邮件威胁的安全解决方案，例如安装反垃圾邮件的软体以提高企业的信息安全。