ChinaVis Data Challenge 2018

**挑战一 评审指南**

本文档针对挑战一给出了详细的评审指南，涵盖了作品提交、问题背景、数据说明、提交给参赛者的任务、潜在答案以及数据中支持这些答案的证据。挑战赛有关问题的完整说明以及向参与者提供的数据，请访问http://www.chinavis.org/2018/challenge.html。

# 一、提交要求

作品提交要求：

（1）答卷：要求参赛者在完成对数据的可视分析后，用图文并茂的方式准确地回答官方预设的问题，以Word或PDF格式提交；

（2）视频：要求参赛者制作带解说的视频，用于解释其可视分析流程，以WMV格式提交，视频总长度不超过5分钟，视频数量1个；

（3）论文：要求参赛者以论文形式，总结其可视分析方案的特点，论文格式要求与ChinaVis论文格式要求一致，篇幅不超过两页，以Word或PDF格式提交；

# 二、背景说明

HighTech是一家互联网高科技公司，有几百名员工，分属财务、人力资源和研发三个部门。公司正在全力研发一款重量级新产品，近期该产品临近发布，公司对内部发生的一切异常现象都非常敏感。为了维护公司的核心利益，确保新产品顺利发布，公司高层决定临时成立内部威胁情报分析小组，该小组将根据公司内部采集到的数据，分析并处置可能存在的各种安全威胁。在分析威胁情报过程中，数据的复杂性需要计算机的处理，但发现、分析与处置安全威胁需要人的经验、认知与判断，可视分析技术能将计算智能与人类智慧这两者紧密结合，通过结合可视化、数据挖掘与人机交互技术，成为威胁情报人员高效分析和理解威胁情报数据的利器。假设您是威胁情报分析小组的成员，请您设计并实现一套可视分析解决方案，帮助该公司及时准确地找出可能存在的内部威胁情报。

# 三、数据支持

给参赛者提供了以下数据支持：

员工登录日志、员工网页访问日志、TCPLOG日志、邮件日志、员工打卡日志。

**登录日志：**员工通过自己主机或跳板机的应用程序，登录服务器或数据库时生成的日志。例如使用SSH、SCP命令、XSHELL程序或者SFTP传输文件都会产生远程登录日志；客户端应用程序访问数据库时，会产生数据库登录日志。

|  |  |  |
| --- | --- | --- |
| **login.csv** | | |
| **字段名称** | **字段含义** | **相关说明** |
| time | 日志生成时间 |  |
| user | 用户名 | 登录使用的用户名 |
| proto | 应用的协议 | 例如ssh、mysql等 |
| dip | 目的IP | 被登录IP |
| dport | 目的端口 | 被登录端口 |
| sip | 源IP | 登录发起IP |
| sport | 源端口 | 登录发起端口 |
| state | 登录结果 | 成功或者失败 |

**网页访问日志：**该日志记录了公司内部所有员工的网页访问记录。time是该条记录生成时间，sip是客户端IP，sport是客户端端口，dip是服务器IP，dport是服务器端口，host是服务器域名。如果通过IP地址直接访问网站，不需要DNS服务器解析，HTTP报头的host字段为空字符串。

|  |  |  |
| --- | --- | --- |
| **weblog.csv** | | |
| **字段名称** | **字段含义** | **相关说明** |
| time | 日志生成时间 |  |
| sip | 源IP | 客户端IP |
| sport | 源端口 | 客户端应用端口 |
| dip | 目的IP | 服务端IP |
| dport | 目的端口 | 服务端应用端口 |
| host | 请求的域名 | HTTP报头的host字段 |

**TCPLOG日志：**记录公司内部网络活动产生的TCP连接。stime、dtime分别是连接建立和断开时间。proto是IP包头中的协议字段值。sip、dip分别是连接发起者和接受者的IP地址，sport、dport是与之对应的源与目的端口。整个连接过程中，sip向dip发送的总字节数为uplink\_length，downlink\_length与之相反。员工的登录行为、网页访问行行为、邮件发送或者接收行为等都会产生一条或者多条TCPLOG日志。

|  |  |  |
| --- | --- | --- |
| **tcpLog.csv** | | |
| **字段名称** | **字段含义** | **相关说明** |
| stime | TCP数据流开始时间 | TCP流的开始时间，即收到该流的第一个SYN包的时间 |
| dtime | TCP数据流结束的时间 | TCP流的结束时间，即收到该流的最后一个包的时间 |
| proto | 协议 | IP包头中的协议字段值 |
| dip | 目的IP | TCP 数据流的服务端IP |
| dport | 目的端口 | TCP 数据流的服务端应用端口 |
| sip | 源IP | TCP 数据流的客户端发起IP |
| sport | 源端口 | TCP 数据流的客户端应用端口 |
| uplink\_length | 上行字节数 | 从TCP流的建立到该流的结束，从客户端发往服务器端的应用层数据的字节总数 |
| downlink\_length | 下行字节数 | 从TCP流的建立到该流的结束，从服务器端发往客户端的应用层数据的字节总数 |

**邮件日志：**邮件日志记录了经过公司邮件服务器的收发邮件信息。time是邮件的发送时间/接收时间，proto是邮件使用的应用协议。sip、dip分别是连接发起者和接受者的IP地址，sport、dport是与之对应的源与目的端口。from、to分别是邮件的发送者和接收者。邮件内容属于隐私，只提供邮件主题subject。

|  |  |  |
| --- | --- | --- |
| **email.csv** | | |
| **字段名称** | **字段含义** | **相关说明** |
| time | 邮件发送/接收时间 | 邮件包头中的邮件发送/接收时间 |
| proto | 应用协议 | SMTP |
| sip | 源IP | IP报头源IP地址 |
| sport | 源端口 | TCP报头源应用端口 |
| dip | 目的IP | IP报头目的IP地址 |
| dport | 目的端口 | TCP报头目的应用端口 |
| from | 邮件发送人 | 来自于邮件头相应字段 |
| to | 邮件接收人 | 来自于邮件头相应字段 |
| subject | 主题 | 来自于邮件头相应字段 |

**打卡日志**：记录了公司每个员工每天上下班时间，一行记录中checkin或者checkout都为0，表示没来上班。那就是说，没有来公司打卡的员工，也有一行考勤记录。另外，如果公司员工当天没来公司上班，则次日该员工会收到旷工提醒邮件。

|  |  |  |
| --- | --- | --- |
| **checking.csv** | | |
| **字段名称** | **字段含义** | **相关说明** |
| id | 员工id |  |
| day | 日期 |  |
| checkin | 上班签到时间 |  |
| checkout | 下班签退时间 |  |

# 四、事件真相

## 4.1 事件简表

|  |  |  |  |
| --- | --- | --- | --- |
| **事件简称** | **重要程度** | **发生时间** | **事件概要** |
| Playing ball 1 | 一般 | 2017-11-02 19点 | 员工集体参加打球活动 |
| Abnormal login 1 | 非常重要 | 2017-11-03 | 登录记录中某账号频繁登录失败 |
| Abnormal login 2 | 非常重要 | 2017-11-04 | 登录记录中某账号频繁登录失败 |
| Abnormal login 3 | 非常重要 | 2017-11-06 | 登录记录中某账号频繁登录失败，后面再没有出现类似事件 |
| Playing ball 2 | 一般 | 2017-11-09 19点 | 员工集体参加打球活动 |
| Playing ball 3 | 一般 | 2017-11-16 19点 | 员工集体参加打球活动 |
| Database failure | 重要 | 2017-11-16 19:22 | 数据库由于员工误操作发生故障 |
| Database maintenance | 重要 | 2017-11-16 19点~23点 | 数据库故障后的临时维护 |
| Snooping product information | 重要 | 2017-11-16 20:22 | 员工在服务器上违规查看产品信息 |
| Playing ball 4 | 一般 | 2017-11-23 19点 | 员工集体参加打球活动 |
| Data leakage | 非常重要 | 2017-11-24 12:43~12:44 | 员工盗取数据并向外泄露 |
| Travel planning | 一般 | 2017-11-27至2017-11-30 | 员工请假旅游 |
| Dimission | 重要 | 2017-11-27至2017-11-28 | 员工申请辞职 |
| Playing ball 5 | 一般 | 2017-11-30 19点 | 员工集体参加打球活动 |
| Financial department overtime | 一般 | 2017-11-19、2017-11-25、2017-11-26 | 月底的周末财务部门多次加班 |
| Stepping stone event | 重要 | 2017-11-17、2017-11-21、2017-11-27、2017-11-30 | 员工通过跳板机向外界服务器上传数据 |
| VPN remote access | 一般 | 2017-11-04、2017-11-05  2017-11-11、2017-11-12  2017-11-18、2017-11-19  2017-11-25、2017-11-26、2017-11-28 | 员工没来公司，通过VPN远程链接到公司内网进行工作 |
| TCPLog system failure | 一般 | 2017-11-10至2017-11-28 | TCP流量监控系统可能存在bug，导致TCP日志数据中部分邮件收发记录的网络协议类型为http，而邮件日志数据中对应记录的协议类型为smtp |

## 4.2 重要人物

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **人物简称** | **员工id** | **部门** | **IP** | **参与事件** |
| Thief | 1487 | 研发3 | 10.64.105.4 | Playing ball 3，Database maintenance，Snooping product information，Data leakage，Dimission |
| Leader1 | 1080 | 研发3 | 10.64.105.175 | Abnormal login 1 |
| Leader2 | 1211 | 研发3 | 10.64.105.63 | Abnormal login 2 |
| Leader3 | 1228 | 研发3 | 10.64.105.146 | Abnormal login 3 |
| DB deleter | 1376 | 研发3 | 10.64.105.219 | Playing ball 3，Database failure，Database maintenance，Dimission |
| DB maintainer | 1284 | 研发3 | 10.64.105.95 | Database maintenance |
| Job leaver | 1281 | 研发2 | 10.64.105.44 | Dimission |
| Traveler1 | 1149 | 人力资源 | 10.64.106.11 | Travel planning |
| Traveler2 | 1352 | 研发3 | 10.64.105.174 | Travel planning |
| Traveler3 | 1383 | 研发3 | 10.64.105.60 | Travel planning |
| Traveler4 | 1389 | 研发3 | 10.64.105.79 | Travel planning |

## 4.2 主线事件

Thief在公司新产品发布前夕盗取产品相关资料并泄露出去。

### 4.2.1 背景和动机

X公司与hightech公司是两个互联网公司，商业竞争极其激烈。X公司为了及时掌握hightech公司的动向，在竞争中取得优势，派遣了员工Thief在hightech公司充当间谍，以获取hightech公司的重要信息并及时反馈回来。Thief在hightech公司工作一年多了，是一名普通的员工。

近期，hightech公司一项重量级新产品临近发布。X公司为了破坏hightech公司的产品发布会，命令Thief在产品发布前夕盗取这一产品的产品信息和相关材料，并在产品发布前泄露出去，以达到打击hightech公司的目的。

Thief了解到产品信息存放在公司的服务器A上。公司有一套完善的内部监控系统，能够将员工的网络活动记录下来。Thief感到非常困惑，如何才能够在不被发现的情况下将数据盗取出来呢？

### 4.2.2 Thief的计划

Thief为了掩饰自己的行为，也为了获取更高的资料查询权限，他决定尝试盗取一个领导的账号，使用这个账号进行盗取数据的操作。Thief先后尝试盗取Leader1、Leader2、Leader3的账号，由于Leader3账号密码设置为弱口令，最终成功破解Leader3的账号密码。

Thief认为使用别人的账号仍然不够保险。后来，他想到了使用跳板机隐藏自己的IP。首先使用Leader3的账号，登录另一台服务器B（10.50.50.43），然后在B服务器上再访问目标服务器A（10.50.50.44），将A服务器上的产品信息等上传到公司外部网络的某台服务器上。盗取数据成功后，Thief决定在这个月月底辞职。

### 4.2.3 Thief的活动

Abnormal login 1：2017-11-03，在这一天的各个时间段，Thief使用Leader1的账号登录A服务器多次，频繁登录失败，破解账号失败。

Abnormal login 2：2017-11-04，在这一天的各个时间段，Thief使用Leader2的账号登录A服务器多次，频繁登录失败，破解账号失败。

Abnormal login 3：2017-11-06，在这一天的各个时间段，Thief使用Leader3的账号登录A服务器多次，频繁登录失败，最终于19:42登录成功，成功破解该账号密码。

Playing ball 3：2017-11-16，Thief报名参加打球活动，由于出现数据库突发故障，事实上并没有参加打球活动，而是参与了数据库维护。

Database maintenance：2017-11-16 19点到23点期间，Thief参与数据库故障后的维护工作，于23:34打卡离开公司。

Snooping product information：2017-11-16 20:22，Thief在维护数据库过程中，使用Leader3的账号登录A服务器，查看服务器上的产品信息，确认A服务器上有他需要的资料。

Data leakage：2017-11-24 12:43-12:44，午休期间，Thief使用Leader3的账号登录B服务器，再使用B服务器做跳板，同时使用Leader3的账号登录A服务器，A服务器向外界服务器泄露数据。

Dimission：2017-11-27提交辞职申请，2017-11-28辞职申请通过，28号之后没有再来公司。

## 4.3 支线事件

### 4.3.1 数据库故障

DB deleter有跳槽的打算，在这个月经常浏览工作招聘类网站。2017-11-16，工作过程中粗心大意不小心误操作了数据库，数据库无法正常工作了。数据库误操作后，系统向Thief和DB maintainer发送了数据库报警邮件。三人当天晚上一起对数据库进行维护，23点半左右维护好数据库，才离开公司。月底，DB deleter辞职离开公司。

Database failure：2017-11-16 19:22，DB deleter误操作了数据库，数据库无法正常工作了，系统向Thief和DB maintainer发送了数据库报警邮件。

Playing ball 3：2017-11-16，Thief和DB deleter都报名参加了打球活动，由于数据库故障，两人都没有参与该活动。

Database maintenance：2017-11-16 19点到23点期间，Thief、DB maintainer和DB deleter参与数据库故障后的维护工作。23点半左右数据库维护成功，Thief于23:34打卡离开公司，DB maintainer于23:35打卡离开公司，DB deleter于23:46打卡离开公司。

### 4.3.2 离职事件

Dimission ：2017-11-27，月底，员工Thief、DB deleter、Job leaver提交了辞职申请，2017-10-28辞职申请审批通过，28号之后这三位员工没有再来公司。DB deleter离职原因是自己早已打算换工作，又遇到数据库故障事件，因此选择月底离职。Job leaver离职原因是家中出现重大变故，突然提出辞职申请。Thief离职原因则是已经完成资料盗取任务，在月底离职集中审批阶段与其他两人一起离职。

### 4.3.3 旅游事件

Travel planning：Traveler1、Traveler2、Traveler3、Traveler4四名员工计划一起出去旅游，他们在2017-11-20至2017-11-24期间频繁浏览旅游网站，在2017-11-24（周五）向各自的领导发送了请假邮件，请假时间为四天（27、28、29、30）。

### 4.3.4 员工集体活动

Playing ball 1-5：2017-11-02、2017-11-09、2017-11-16、2017-11-23，2017-11-30，这五天（都是周四）早上9:30 hr（hr@hightech.com）向所有员工（allstaff@hightech.com）发送邮件“打球啦，欢迎大家参加”，有意向参加的员工回复确认邮件，并在19:30前离开公司参加活动。大部分参加活动的员工均在19:00-19:20这段时间区间离开公司。

### 4.3.5 财务月底加班

Financial department overtime：由于月底财务工作繁忙，在月底的周末，2017-11-19、2017-11-25、2017-11-26三天，财务部门绝大部分的员工来到了公司工作加班。

### 4.3.6 跳板机事件

Stepping stone event：除了泄露数据的员工Thief外，在2017-11-17、2017-11-21、2017-11-27、2017-11-30四天，1183、1273、1169、1151四名员工先后通过跳板机向外界服务器上传数据，但这些都属于正常行为。

### 4.3.7 VPN远程访问

VPN remote access：1147、1283、1284、1328、1334、1376、1487、1494八名员工曾在周末通过VPN远程链接到公司内网加班工作；1059在2017-11-28周二没来公司，通过VPN远程链接到公司内网审批了员工1376和1487的辞职申请。

### 4.3.8 流量监控系统故障

TCPLog system failure：2017-11-10至2017-11-28，TCP日志数据中一些邮件访问日志的协议本应该是smpt，却被标记为了http，可能是由于TCP日志系统bug造成。

# 五、参考答案

评审总体说明：

1：要求回答问题准确且简明扼要；

2：要求以可视分析为主要技术路线探索问题答案；

3：要求用可视化的方式呈现与解释给出的答案；

4：鼓励给出参考答案以外的任何合理的新发现；

5：鼓励在分析过程中引入智能算法；

6：鼓励参赛队伍自行开发新颖的可视分析解决方案；

7：鼓励参赛队伍使用自己团队（公司）研发的分析工具。

## 5.1 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（建议参赛者回答此题文字不多于500字，图片不多于5张）；**

该公司组织结构较为简单清晰，总裁1人，总管全局，5个部门（财务、人力、研发1、研发2、研发3）各有一个部门主管，3个研发分为很多个小组，每个小组有一个组长。具体信息如表:

|  |  |  |  |
| --- | --- | --- | --- |
| **部门** | **人数** | **领导** | **员工（红色为小组组长）** |
| 总裁 | 1 | 1067 |  |
| 财务 | 24 | 1041 | 1368,1347,1255,1248,1327,1439,1137,1370,1467,1226,1369,1186,1213,1451,1124,1431,1293,1253,1342,1498,1108,1180,1346 |
| 人力资源 | 18 | 1013 | 1104,1499,1371,1184,1251,1295,1312,1433,1165,1300,1378,1473,1118,1363,1249,1110,1149 |
| 研发1 | 62 | 1068 | 1154, 1176,1315, 1152,1420;  1191, 1428, 1483, 1469, 1156, 1456, 1204, 1435;  1207, 1189, 1330, 1319, 1296, 1399, 1263, 1103;  1100, 1139, 1481, 1385, 1147, 1321, 1493, 1458, 1170, 1379, 1305, 1234,1362, 1405, 1159, 1474;  1098, 1343, 1127, 1496, 1277, 1334;  1209, 1460, 1126, 1322, 1339, 1388, 1349, 1153;  1060, 1359, 1457, 1328, 1145, 1306, 1440, 1396, 1446, 1336; |
| 研发2 | 88 | 1007 | 1087, 1151, 1220, 1286, 1141, 1494, 1373;  1115, 1233, 1423, 1471, 1243, 1491, 1464, 1169, 1408, 1183, 1425, 1357, 1459, 1455;  1230, 1167, 1182, 1354, 1265, 1129, 1252, 1223, 1404, 1200;  1172, 1132, 1490, 1246, 1466, 1475, 1314, 1397, 1436, 1480, 1257, 1345, 1477;  1192, 1282, 1403, 1303, 1210, 1340, 1140, 1484;  1199, 1348, 1391, 1278, 1197, 1486;  1092, 1270, 1344, 1112, 1308, 1301;  1125, 1307, 1398, 1113;  1224, 1281, 1275, 1406, 1323, 1102, 1299, 1134, 1326, 1106, 1416, 1205, 1195, 1221, 1495, 1393, 1429, 1351, 1417; |
| 研发3 | 106 | 1059 | 1080, 1364, 1181, 1449, 1311, 1193, 1422, 1194, 1297, 1384, 1376;  1211, 1411, 1287, 1382, 1231, 1365, 1284, 1497, 1164;  1101, 1356, 1241, 1461, 1313, 1352, 1175, 1350, 1179, 1338, 1325;  1143, 1434, 1380, 1438, 1367, 1355, 1279, 1163, 1324, 1304, 1381, 1217;  1119, 1135, 1238, 1244, 1268, 1401, 1148, 1274, 1360, 1390, 1291;  1155, 1421, 1216, 1470, 1409, 1462, 1444, 1332, 1206, 1283, 1389, 1267;  1058, 1261, 1171, 1333, 1424, 1445, 1450, 1202, 1130, 1383, 1245, 1489;  1228, 1290, 1465, 1178, 1177, 1174, 1394, 1487, 1273;  1096, 1402, 1478, 1239, 1500, 1254;  1079, 1262, 1395, 1219, 1482;  1057, 1173, 1374, 1410, 1361, 1150, 1142; |

## 5.2 挑战1.2

**分析该公司员工的日常工作行为，按部门总结员工的正常工作模式（建议参赛者回答此题文字不多于1000字，图片不多于8张）；**

建议按部门来讨论员工的工作行为，5个部门在工作时间、常用服务器和常用网站上有一定区别，也可以从邮件主题分析中了解到各部门工作重点略有不同，具体见下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **部门** | **工作时间** | **常用服务器** | **常用网站（访问量按降序排列）** | **说明** |
| 财务 | 8点~17点 | 10.63.120.70（OA）、10.5.71.60（Email） | email.hightech.com、  OA.hightech.com、www.baidu.com、[www.google.com、](http://www.google.com、)  ju.taobao.com、[www.so.com、](http://www.so.com、)  [www.bankcomm.com、ai.taobao.com、store.apple.com](http://www.bankcomm.com、ai.taobao.com、store.apple.com)、ent.163.com | 工作内容主要与财务有关 |
| 人力资源 | 9点~18点 | 10.63.120.70 （OA）、  10.5.71.60 （Email） | email.hightech.com、  OA.hightech.com、  [www.google.com、](http://www.google.com、)  [www.yahoo.com、](http://www.yahoo.com、)  [www.baidu.com、](http://www.baidu.com、)  ai.taobao.com、  [www.ccb.com、](http://www.ccb.com、)  china.alibaba.com、ju.taobao.com、  www.baihe.com | 工作内容主要包括考勤、绩效考核、福利保障、员工招聘等 |
| 研发1 | 9点~18点 | 频繁访问下列服务器：10.5.71.60（Email）、10.63.120.70（OA）、10.50.50.26（git）、10.50.50.27（jira）、10.50.50.28（lib01）、10.50.50.29（lib02）  此外还访问下列服务器：10.7.133.15、10.7.133.16、10.7.133.21、10.7.133.22、10.50.50.30、10.50.50.31、10.50.50.33、10.50.50.35、10.50.50.36、10.50.50.37、10.50.50.38、10.50.50.40、10.50.50.41、10.50.50.43、10.50.50.44、10.50.50.45、10.50.50.46、10.50.50.48 | email.hightech.com、git.hightech.com、OA.hightech.com、jira.hightech.com、lib01.hightech.com、lib02.hightech.com、www.ruanyifeng.com、www.baidu.com、www.tianya.cn、www.csdn.net | 工作内容主要包括研发工作、技术分享等。三个研发部门各自工作重点没有明显区别 |
| 研发2 | 9点~18点 | 频繁访问下列服务器：10.5.71.60（Email）、10.63.120.70（OA）、10.50.50.26（git）、10.50.50.27（jira）、10.50.50.28（lib01）、10.50.50.29（lib02）  此外还访问下列服务器：10.7.133.15、10.7.133.16、10.7.133.18、10.7.133.19、10.7.133.20、10.50.50.33、10.50.50.37、10.50.50.38、10.50.50.40、10.50.50.43、10.50.50.46、10.50.50.48、10.50.50.49 | email.hightech.com、git.hightech.com、OA.hightech.com、jira.hightech.com、lib01.hightech.com、lib02.hightech.com、[www.baidu.com、www.programmer.com.cn、www.ruanyifeng.com](http://www.baidu.com、www.programmer.com.cn、www.ruanyifeng.com)、www.yahoo.com | 工作内容主要包括研发工作、技术分享等。三个研发部门各自工作重点没有明显区别 |
| 研发3 | 10点~19点 | 频繁访问下列服务器：10.5.71.60（Email）、10.63.120.70（OA）、10.50.50.26（git）、10.50.50.27（jira）、10.50.50.28（lib01）、10.50.50.29（lib02）  此外还访问下列服务器：10.7.133.16、10.7.133.19、10.7.133.20、10.50.50.31、10.50.50.33、10.50.50.34、10.50.50.36、10.50.50.37、10.50.50.38、10.50.50.39、10.50.50.40、10.50.50.41、10.50.50.42、10.50.50.43、10.50.50.44、10.50.50.46、10.50.50.47、10.50.50.48、10.50.50.49 | email.hightech.com、git.hightech.com、OA.hightech.com、jira.hightech.com、lib01.hightech.com、lib02.hightech.com、www.baidu.com、www.google.com、www.programmer.com.cn、www.ruanyifeng.com | 工作内容主要包括研发工作、技术分享等。三个研发部门各自工作重点没有明显区别 |

其它模式说明：

1. 午间12:30-13:30为午休时间，这段时间员工一般进行午餐、小憩或者浏览网页，大部分员工有午间时间浏览网页的习惯。
2. 只有研发1、研发2、研发3三个部门有服务器登录操作。
3. 各部门领导工作时间比较弹性，有较多迟到、旷工现象，但属于正常情况。
4. 财务部的邮件内容主要与财务分析、资金、会计核算、税务、成本控制、财务报销等有关；人力资源部的邮件内容主要与员工招聘、劳动合同、考勤、绩效考核、福利保障等有关；研发部的邮件内容主要与需求分析、软件开发等产品研发类的名词有关。

## 5.3 挑战1.3

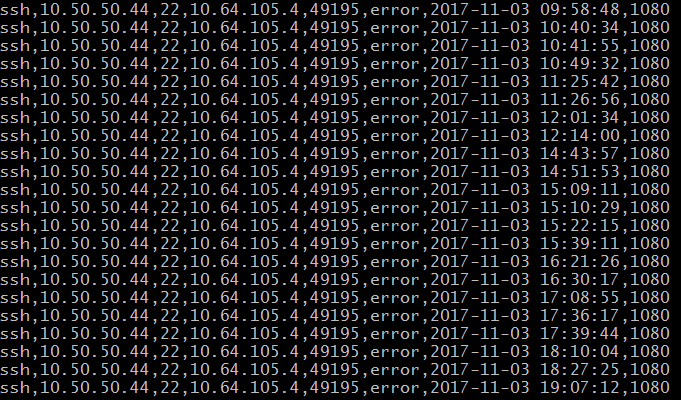
**找出至少5个异常事件，并分析这些事件之间可能存在的关联，总结你认为有价值的威胁情报（建议参赛者回答此题文字不多于1500字，图片不多于10张）。**

4.1事件简表中所有18个事件都可以看作为异常事件。为了更好的叙述整个故事，我们推荐将4.1事件简表中所有事件进行分类聚合。在参考答案中，我们将事件简表中所有事件聚合为9个聚合事件，分别是：数据泄露、数据库故障、员工离职、跳板机事件、员工请假旅游、员工集体活动、财务月底加班、VPN远程访问、流量监控系统故障，具体见后续事件说明。

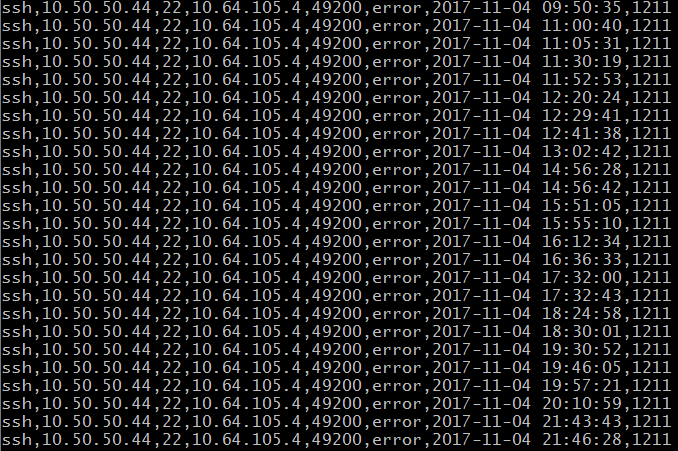
关联关系分析我们推荐从两个角度来做，一个角度是时间线分析；另一个角度是情报主体分析，情报主体可以是员工、客户端、服务器、邮件等数据中可以抽象出的实体或对象，这些对象之间的复杂关系构成了情报线索，比如：2017年11月16日，群发打球通知邮件，推测Playing ball 3事件的发生；当天员工Thief报名参加打球但是最后没有去，原因是当天傍晚收到了数据库故障邮件，推测员工Thief与数据库故障有关系；员工1228当天晚上访问了有产品重要信息的服务器，但是所用IP并不是员工1228常用的客户端IP，推测这个访问与早些天发生的Abnormal login事件有联系。本题我们希望作品最好能够用情报分析思路，通过情报主体的抽象，分析和挖掘这些主体间关联，来完成整个分析过程。

### 5.3.1 数据泄露

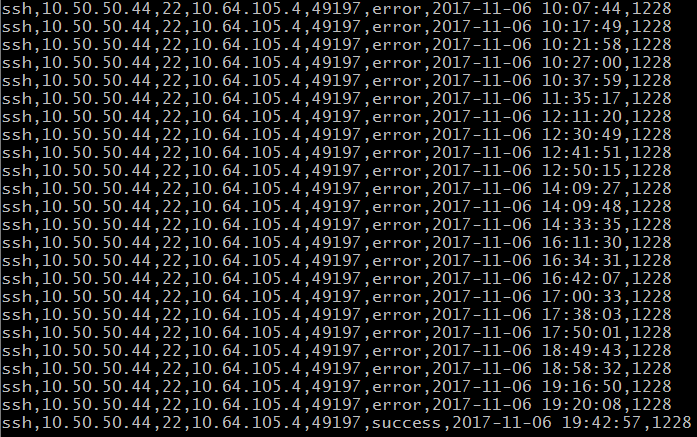
Abnormal login 1：2017-11-03，在这一天的各个时间段，员工1487（IP为10.64.105.4）使用账号1080登录服务器10.50.50.44多次，频繁登录失败，破解账号失败。典型数据记录如下，摘录自2017-11-03的login.csv日志文件：



Abnormal login 2：2017-11-04，在这一天的各个时间段，员工1487（IP为10.64.105.4）使用账号1211登录服务器10.50.50.44多次，频繁登录失败，破解账号失败。典型数据记录如下，摘录自2017-11-04的login.csv日志文件：



Abnormal login 3：2017-11-06，在这一天的各个时间段，员工1487（IP为10.64.105.4）使用账号1228登录服务器10.50.50.44多次，频繁登录失败，最终于19:42登录成功，成功破解该账号密码。典型数据记录如下，摘录自2017-11-06的login.csv日志文件：



Snooping product information：2017-11-16 20:22，员工1487（IP为10.64.105.4）使用账号1228登录服务器10.50.50.44，查看服务器上的产品信息。典型数据记录如下，摘录自2017-11-16的login.csv日志文件：



Data leakage：2017-11-24 12:43~12:44，员工1487（IP为10.64.105.4）使用账号1228登录服务器10.50.50.43，再使用服务器10.50.50.43做跳板，同时使用账号1228登录服务器10.50.50.44。典型数据记录如下，摘录自2017-11-24的login.csv日志文件：

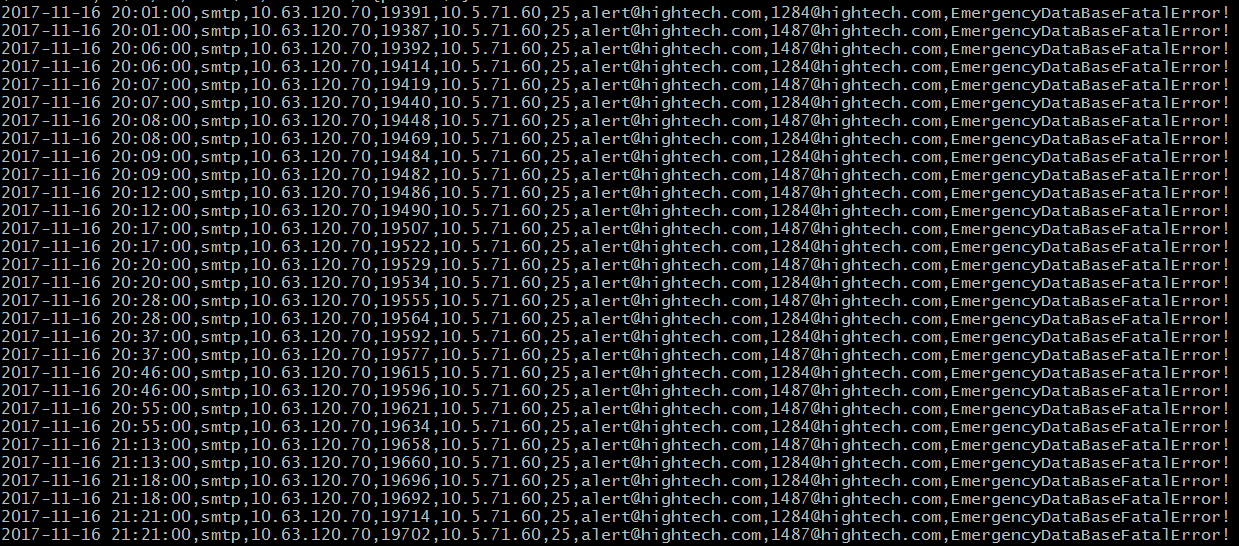


再查看服务器10.64.105.4、10.50.50.43、10.50.50.44在2017-11-24 12:43~12:44左右的TCP记录。10.64.105.4（1487对应的IP）使用ssh协议访问服务器10.50.50.43，服务器10.50.50.43使用ssh协议访问服务器10.50.50.44，服务器10.50.50.44再使用ssh协议访问了13.250.177.223（外部不明服务器），发现有较大的流量传输，上传了600M左右的数据。典型数据记录如下，摘录自2017-11-24的tcpLog.csv日志文件：



### 5.3.2 数据库故障

Database failure：2017-11-16 20点之后，在员工1487（10.64.105.4）和1284（10.64.105.95）的邮件记录中发现大量邮件“EmergencyDataBaseFatalError”，可能是数据库发生了故障。典型数据记录如下，摘录自2017-11-16的email.csv日志文件：

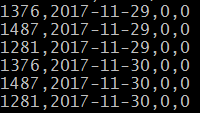


Database maintenance：员工1487、1376、1284在19点到23点期间频繁访问10.63.120.70（OA）服务器，原因是服务器10.63.120.70上的数据库出了故障。于是这三位员工在19点到23点期间维修该服务器上的数据库。23点半左右维护好数据库，他们才离开公司。

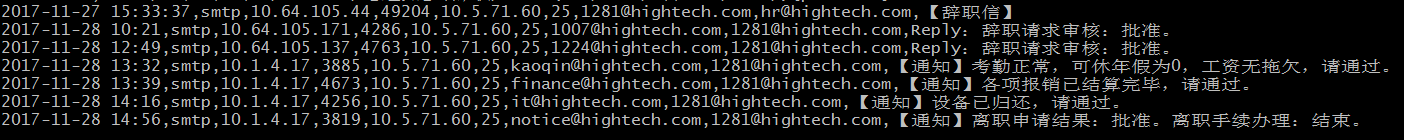
### 5.3.3 员工离职

Dimission：从考勤记录和邮件记录发现，2017-11-27，员工1376、1487、1281递交辞职信申请辞职，2017-11-28申请审核通过，28号之后这三位员工没有再来公司。

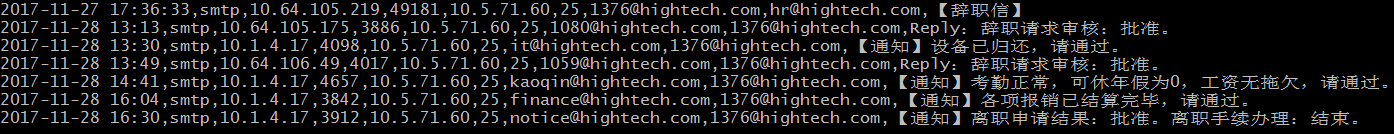
2017-11-29至2017-11-30员工1376、1487、1281没有来公司。典型数据记录如下，摘录自2017-11-29和2017-11-30的checking.csv日志文件：



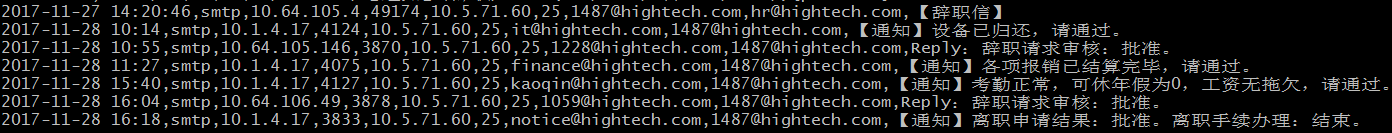
员工1281的辞职过程如下，2017-11-27递交“辞职信”，2017-11-28辞职申请审核通过。典型数据记录摘录自2017-11-27和2017-11-28的email.csv日志文件：



员工1376的辞职过程如下，2017-11-27递交“辞职信”，2017-11-28辞职申请审核通过。典型数据记录摘录自2017-11-27和2017-11-28的email.csv日志文件：



员工1487的辞职过程如下，2017-11-27递交“辞职信”，2017-11-28辞职申请审核通过。典型数据记录摘录自2017-11-27和2017-11-28的email.csv日志文件：



### 5.3.7 跳板机事件

Stepping stone event：2017-11-17、2017-11-21、2017-11-27、2017-11-30四天，1183、1273、1169、1151四名员工先后通过两次跳板向外界服务器13.250.177.223上传数据。

2017-11-17，14:49，1183 从自己的客户端10.64.105.165登录到10.7.133.20，再从此IP登录到10.50.50.40，向13.250.177.223上传数据。典型数据记录如下，分别摘录自2017-11-17的login.csv日志文件和tcpLog.csv日志文件：





2017-11-21，13:31，1273从自己的客户端10.64.105.244登录到10.50.50.49，再从此IP登录到10.50.50.34，向13.250.177.223 上传数据。典型数据记录如下，分别摘录自2017-11-21的login.csv日志文件和tcpLog.csv日志文件：





2017-11-27，21:03，1169 从自己的客户端10.64.105.199登录到10.50.50.37，再从此IP登录到10.50.50.46，向 13.250.177.223上传数据。典型数据记录如下，分别摘录自2017-11-27的login.csv日志文件和tcpLog.csv日志文件：





2017-11-30，17:19，1151从自己的客户端10.64.105.73登录到10.50.50.49，再从此IP登录到10.7.133.16，向13.250.177.223上传数据。典型数据记录如下，分别摘录自2017-11-30的login.csv日志文件和tcpLog.csv日志文件：

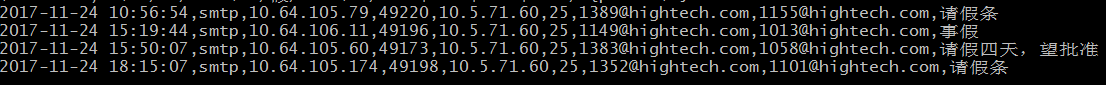




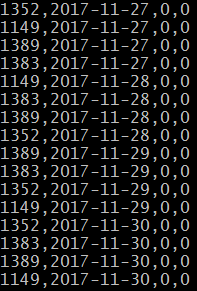
### 5.3.4 员工请假旅游

Travel planning：从考勤记录和邮件记录发现员工1149、1352、1383、1389都在2017-11-24（周五）请假，在2017-11-27至2017-11-30期间没来公司，请假时间跨度相同。结合网页访问记录发现这些员工于2017-11-20至2017-11-24期间频繁访问旅游类网站，可能4人一起出去旅游。

2017-11-24，四人向各自的领导发送请假邮件。典型数据记录如下，摘录自2017-11-24的email.csv日志文件：



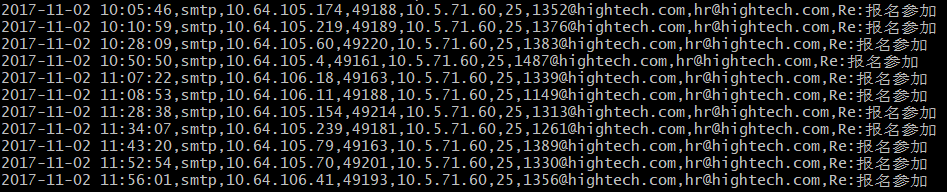
2017-11-27至2017-11-30期间没来公司。典型数据记录如下，摘录自2017-11-27至2017-11-30的checking.csv日志文件：



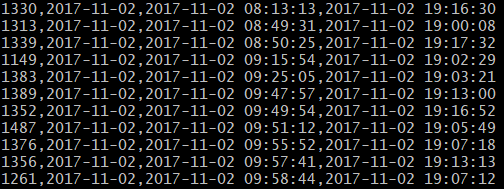
### 5.3.5 员工集体活动

2017-11-02、2017-11-09、2017-11-16、2017-11-23，2017-11-30，这五天（都是周四）早上9:30 hr（hr@hightech.com）向所有员工（allstaff@hightech.com）发送邮件“打球啦，欢迎大家参加”，有意向参加的员工回复确认邮件，并在19:30前离开公司参加活动。

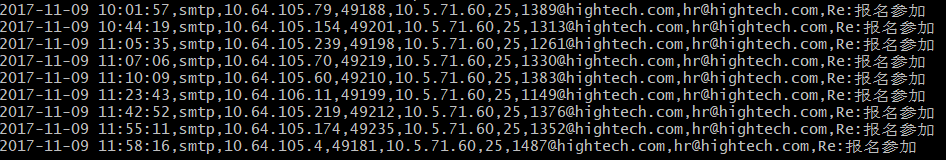
Playing ball 1：2017-11-02，员工1352、1376、1383、1487、1339、1149、1313、1261、1389、1330、1356报名参加了打球活动。典型数据记录如下，摘录自2017-11-02的email.csv日志文件：



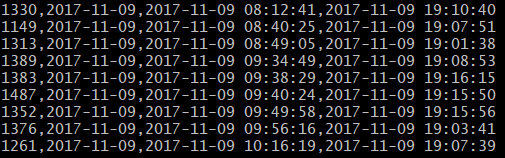
报名的员工在19:00~19:20这段时间区间离开公司参加打球活动。典型数据记录如下，摘录自2017-11-02的checking.csv日志文件：



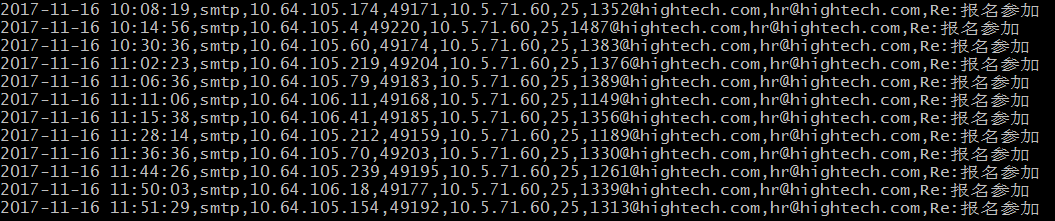
Playing ball 2：2017-11-09，员工1389、1313、1261、1330、1383、1149、1376、1352、1487报名参加了打球活动。典型数据记录如下，摘录自2017-11-09的email.csv日志文件：



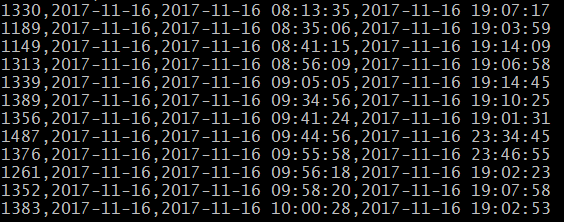
报名的员工在19:00~19:20这段时间区间离开公司参加打球活动。典型数据记录如下，摘录自2017-11-09的checking.csv日志文件：



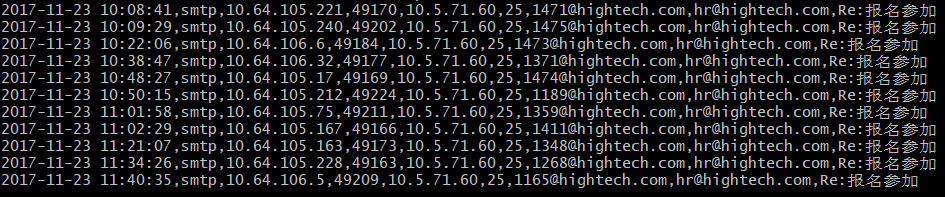
Playing ball 3：2017-11-16，员工1352、1487、1383、1376、1389、1149、1356、1189、1330、1261、1339、1313报名参加了打球活动。典型数据记录如下，摘录自2017-11-16的email.csv日志文件：



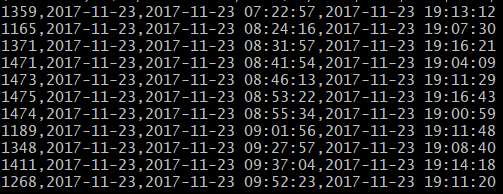
参加打球活动的员工在19:00-19:20期间打卡离开公司，1487和1376报名参加打球活动，但从打卡记录来看没有参与。典型数据记录如下，摘录自2017-11-16的checking.csv日志文件：



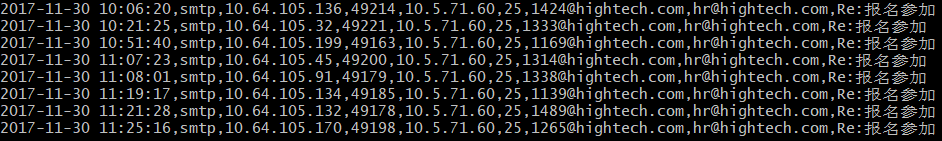
Playing ball 4：2017-11-23，员工1471、1475、1473、1371、1474、1189、1359、1411、1348、1268、1165报名参加了打球活动。典型数据记录如下，摘录自2017-11-23的email.csv日志文件：



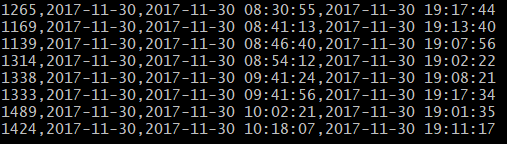
报名的员工在19:00~19:20这段时间区间离开公司参加打球活动。典型数据记录如下，摘录自2017-11-23的checking.csv日志文件：



Playing ball 5：2017-11-30，员工1424、1333、1169、1314、1338、1139、1489、1265报名参加了打球活动。典型数据记录如下，摘录自2017-11-30的email.csv日志文件：



报名的员工在19:00~19:20这段时间区间离开公司参加打球活动。典型数据记录如下，摘录自2017-11-30的checking.csv日志文件：



### 5.3.6 财务月底加班

Financial department overtime：2017-11-19、2017-11-25、2017-11-26三天，财务部门绝大部分的员工来到了公司工作，其它部门没有发生该现象。财务部门加班的数据统计情况如下：

|  |  |  |  |
| --- | --- | --- | --- |
| **部门** | **总人数** | **日期** | **加班人数** |
| 财务 | 24 | 2017-11-19（周日） | 15 |
| 2017-11-25（周六） | 20 |
| 2017-11-26（周日） | 21 |

### 5.3.8 VPN远程访问

VPN remote access：正常情况下，如果员工当天没有来公司却产生了TCP流量，这是员工通过VPN远程链接公司内网进行工作导致的。数据中出现了员工产生TCP流量却未打卡（无打卡记录或打卡记录的checkin和checkout均为0）的异常现象，这些现象大多发生在周六和周日，涉及人员包括1147、1283、1284、1328、1334、1376、1487、1494、1059九名员工。具体信息如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **日期** | **周** | **工作日/休息日** | **涉及人员** | **事件描述** |
| 2017-11-04 | 周六 | 休息日 | 1487 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-05 | 周日 | 休息日 | 1147、1328、1334、1494 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-11 | 周六 | 休息日 | 1147、1328、1376、1487、1494 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-12 | 周日 | 休息日 | 1376 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-18 | 周六 | 休息日 | 1147、1283、1284、1328、1334、1376、1487、1494 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-19 | 周日 | 休息日 | 1487 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-25 | 周六 | 休息日 | 1283、1284、1376、1487 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-26 | 周日 | 休息日 | 1376、1487 | 员工通过VPN远程链接到公司内网加班工作 |
| 2017-11-28 | 周二 | 工作日 | 1059 | 员工通过VPN远程链接到公司内网审批了自己主管部门的两名员工（1376和1487）的辞职申请 |

### 5.3.9 流量监控系统故障

TCPLog system failure：2017-11-10至2017-11-28，存在部分TCP记录网络协议类型为http，目的端口为25（smtp协议常用端口）的情况，这部分 TCP记录的目的IP均为10.5.71.60（邮件服务器）。进一步检查tcpLog表和email表可以找到对应的邮件记录，但是两表中对应记录的网络协议类型不一致，在email表中网络协议为smtp，在tcpLog表中网络协议为http。该现象是日志记录系统故障导致。

典型错误日志信息如下（tcpLog表和email两表通过源IP、目的IP、时间建立连接条件）

