基于内网数据的威胁情报可视分析

陈明毅，蔡真真，韩瑶鹏，蹇诗婕，田甜

摘要—威胁态势千变万化，为了辅助分析人员从大量异构且复杂的数据中发掘出有效的威胁情报，本文通过对给定数据进行聚类、组合，使用树图、力导向图展现组织结构，使用折线图、玫瑰图、桑基图等展现部门的行为模式，同时辅以词云、平行坐标图、日历图等发现异常事件。利用本文提出的多视图协同可视分析方法，可以全方位多角度发现 HighTech 公司的组织结构、行为模式和异常事件，并总结为有价值的威胁情报。

关键词—多视图协同可视分析、威胁情报、组织结构、行为模式、异常事件

概述

本文首先对所提供的内网数据进行分析，并对数据进行清洗（主要表现为编码转换）。提供的数据主要包含：打卡日志、邮件收发日志、网页浏览日志、登录和流量日志。而后通过数据之间的关联关系，确定“人员-IP地址-邮箱”的三元唯一组合，并通过邮件收发关系及主题确定组织结构。

在可视分析阶段，利用多视图协同分析的方式从不同维度为分析人员提供支持，表现为利用折线图、桑基图、平行坐标图等不同样式的图形，透过打卡、登录及流量日志确定工作模式和发现异常事件。

为了保证用户体验，可视化系统分为三个页面，分别展示组织结构、组织工作模式和个人相关信息。同时为了保证展示速度，优化了各种联合查询算法，并将分析结果缓存，无需重复查找数据库并处理，将结果直接推送到前端页面。下文分点介绍具体的可视化实现。

1 组织结构可视分析

组织结构是分析组织行为模式和异常事件的基石。为了准确的分析出组织结构及人员构成，我们对邮件的收发人及收发主题进行了整理，利用算法剔除垃圾邮件，对剩下的邮件主题，利用jieba分词库，进行分词聚类，并结合常识认知，可以得到如下图1.1所示的员工组织结构树图。

图 1. 员工组织结构图（树图）

如图1所示，HighTech公司由管理层领导，下设三大部门，每个部门用不同颜色表示。紫色区域为管理层，由1067担任总经理；黄色代表人力资源部门，由1013担任部长；橙色代表财务部门，由1041担任部长；红色代表研发部门，其中研发部门又下设三个小部门（后文以研发部门1等区分三个小部门），每个小部门又下设若干个研发小组（后文以研发小组1-1等区分研发小组），每个研发小部门和研发小组均设有一个负责人。

- *陈明毅，中国科学院信息工程研究所， chenmingyi@iie.ac.cn*
- *蔡真真，中国科学院信息工程研究所， caizhenzhen@iie.ac.cn*
- *韩瑶鹏，中国科学院信息工程研究所，15152108971@163.com*
- *蹇诗婕，北京科技大学，17888838363@163.com*
- *田　甜，中国科学院信息工程研究所， tiantian@iie.ac.cn*

2 行为模式可视探索

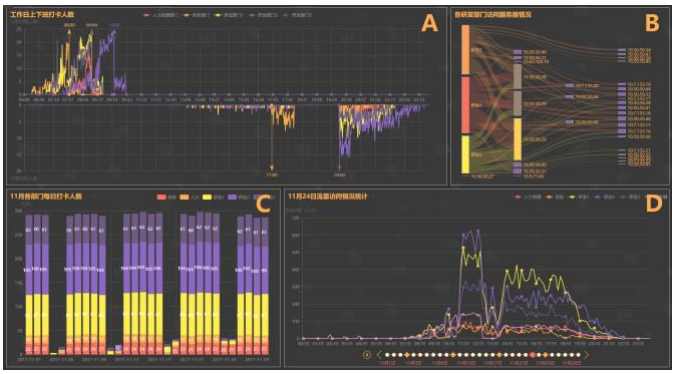

图 2. 部门画像统计页图。A 部分为各部门工作日上下班统计时间折线图；B 部分为内部资产 TCP 联系桑基图；C 图为各部门每日打卡柱状图；D 图为每日各部门流量数量折线图。

2.1 邮件发送模式

分析邮件主题和收发件人，可以发现人力资源部门和财务部门的普通员工会定期向部长发送工作汇报；研发小组普通员工会向组长发送工作总结，研发小组组长会向各研发部门部长发送工作总结；五位部长向总经理进行工作汇报。

2.2 时间模式

折线图适合用于集中表现时间模式。如图 2-A 与图 2-D 所示，通过峰谷值与上升下降趋势可以较为直观的分析出五个部门的上下班时间模式和上班时间安排：

表 1. 各部门上下班时间模式

部门	上班时间	下班时间
人力资源	9:00	20:00
财务	8:00	17:00
研发部门 1	9:00	20:00
研发部门 2	10:00	20:00
研发部门 3	9:00	20:00

表 2. 各部门上班时间安排

时间	人力资源	财务	研发 1	研发 2	研发 3
9:30-10:00	例会	例会			
10:20-10:50			例会	例会	例会
12:30-13:00	午休	午休	午休	午休	午休
13:20-13:50	午休	午休	分享	分享	分享

同时，根据每日打卡人数统计，可以发现 HighTech 公司实行双休制，周末除特殊情况不加班。

2.3 资产模式

HighTech 公司的资产主要分为内部资产和外部资产。内部资产由员工主机和服务器组成。绝大部分服务器是各个研发组混合使用，并且没有专门的用途，同时运行多个服务（尤其是多个类型的数据库）；个别服务器是专门服务于某个研发部门。

外部资产则是 HighTech 公司在公网上部署有多台服务器（或是合作伙伴服务器）。部分研发人员也有自己的服务器，涉及境内外数百台服务器。内部资产会和外部资产之间进行大量的数据传输。桑基图可以将联系与数量组合一起，而如图 2-B 显示的内部资产之间桑基图可以通过数量发现孤立的连接事件。

2.4 其他模式

- 研发部门 3 会给外部人员发送技术分享及探讨邮件；
- 财务在月底较为频繁的加班。

3 异常行为及威胁情报可视分析

在异常行为分析及总结威胁情报部门，我们是通过多视图协同分析的方法，透过异常的行为挖掘出更多的异常行为，最后总结成威胁情报。以下按照威胁情报进行详细阐述。

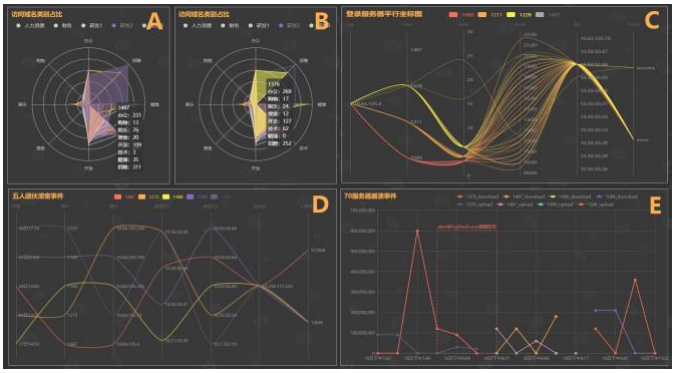

图 3. 威胁情报详情页图。A 部分为员工 1487 访问域名类别占比雷达图；B 部分为员工 1376 访问域名类别占比雷达图；C 部分为 10.64.105.4 员工主机登录服务器平行坐标图；D 部分为员工 1487 等 5 人团体通过内部服务器跳转异常事件平行坐标图；E 图为 10.63.120.70 服务器崩溃事件折线图。

3.1 员工 1487 等 5 人团体具有泄密风险

员工 1487 近一个月有多项异常行为，这些行为指向其是具有窃取公司内部资料风险的异常员工。同时，1183、1273、1169、1151 因与 1487 具有部分相同的异常行为，我们认定这五人是具有泄密风险的团体。

3.1.1 多次尝试使用他人账户登录服务器，并成功登录

平行坐标图适合用于分析多维度数据间的关联关系。分析员工 1487 登录行为的平行坐标图（图 3-C 所示），可以发现其分别在 3、4、6 日频繁尝试 1080、1211 和 1228 三位研发组长账户，最终于 6 日晚成功使用 1228 账户登录 10.50.50.44 服务器。其后又在 16 日夜间和 24 日午间再次盗用 1228 账户成功登录 10.50.50.44 服务器和 10.50.50.43 服务器。

3.1.2 通过内部服务器向境外服务器上传数据

分析 1487 在 24 日盗用账户的异常登录行为，发现其通过 10.50.50.43 做跳板登录 10.50.50.44 服务器，并向境外服务器 13.250.177.223 上载了大量数据。

分析 13.250.177.223 服务器，发现仅有五条访问记录，对其进行追踪溯源，发现 1183，1273，1487，1169 和 1151 五人，均经过两次跳转，向该服务器上载数据。跳转情况和上传数据如图 3-D 所示。这种显而易见的跳转隐匿行为具有极高风险。

3.1.3 大量访问招聘网站，27 号提出离职

对 1487 访问域名进行雷达图分析（图 3-A 所示），发现其大量访访问招聘网站，于 27 号提出离职。

3.2 员工 1376 具有破坏、窃取数据风险

3.2.1 尝试大量下载数据并造成数据库告警

分析 alert 告警邮箱的主题词云，发现其在 16 日晚 20 点向员工 1284、1487 发送了数十封数据库异常告警邮件。分析两位员工登录服务器的平行坐标图，发现其均在 20 点 30 分之后登录 10.63.120.70 服务器进行维修。对事发前后的流量进行可视分析（图 3-E 所示），发现只有 1376 使用 ssh 协议频繁访问服务器并下载大量内容，具有破坏服务器并窃取数据的动机。

3.2.2 大量访问招聘网站，并且与其余两人同天提出离职

1376 与 1487 类似，在近一个月大量访问招聘网站，达 252 次，比例接近 35%（图 3-B 所示）。于 27 号与 1281、1487 同时提出离职。

3.3 员工 1376、1487 等 9 人团体具有潜逃风险

根据邮件主题聚类结果，得知该公司每周举行打球活动。对参与员工做关联分析，发现 1149、1261、1313、1330、1352、1376、1383、1389、1487 这 9 名员工均只参与前三周的打球活动，判定其极有可能是一个小团体。

对该 9 名员工打卡记录和邮件关联关系进行可视分析，发现员工 1149（人力资源部门）、1352（研发小组 2-6）、1383（研发小组 2-2）、1389（研发小组 2-9）该四名员工在 27 日至 30 日四天均旷工，且无共同出差可能。因此猜测 4 名员工极有可能伙同打球团体中已离职的 1376、1487 参与犯罪并潜逃。建议对剩余三名员工 1261、1313、1330 进行调查。

3.4 孤立异常事件

3.4.1 服务器没有按计划访问

服务器 10.50.50.27 在每周六凌晨 3 点固定访问服务器 10.7.133.15，而在月末的 25 日却没有任何访问记录。

3.4.2 公司过量招聘

该公司仅有 299 名员工，但于 11 月发送了 149 封录用通知和 149 封 Offer，招聘人数与公司规模严重不符。

3.4.3 研发组长充当猎头

该公司 20 位研发组长收到累计 257 封来自猎聘网的候选人邮件，职位申请包括政府事务经理、销售经理、组织发展总监等，研发组长均注册猎头账号招人，且申请职位与研发无关。

4 威胁预警与应对措施

根据上述通过可视分析得到的异常事件及总结的威胁情报，可以提出以下的威胁预警及应对措施：

- 员工 1376、员工 1487 具有重大窃取、破坏敏感数据风险，鉴于其已离职，应对其所持有或使用过的资产进行检查，确认是否会产生后续风险；
- 除员工 1376、员工 1487 之外的剩余 7 人具有协助 1376、1487 进行数据窃取的风险，且未辞职或已潜逃，应对其进行约谈，视结果采取进一步措施，防范可能造成的后续风险。