网络安全创新大会
Cyber Security Innovation Summit

基于流量的敏感数据异常访问行为识别方法

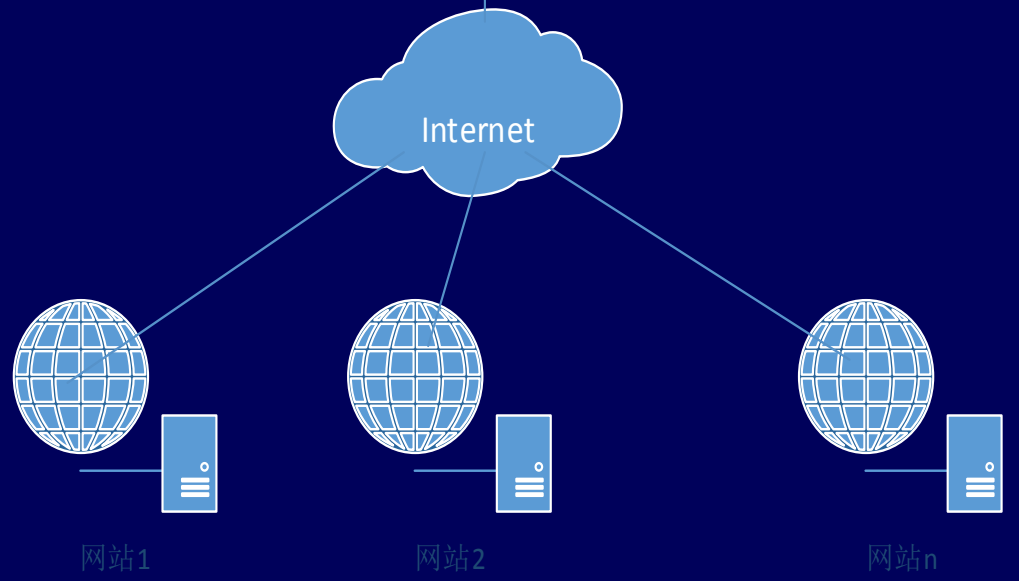
民航面临的数据安全威胁

为建设智慧民航，提升航空公司运行效率，为旅客提供真情服务，近年来民航企事业单位大力推进信息化建设，信息系统的种类和数量均越来越多，信息系统中所承载的数据也愈加重要和敏感。旅客信息作为民航的核心数据之一，近年来相关数据量爆发式增长，但是针对旅客信息泄露的网络安全事件屡屡发生。为保护民航旅客信息安全，维护旅客的合法权益，民航各单位针对旅客信息的保护工作压力越来越大。

数据安全问题

2017年，浙江省温州市公安局网安支队，苍南县公安局网警大队破获一起特大黑客攻击窃取国内航空公司网站信息案件。黑客非法入侵50多家民用航空类公司网站，窃取乘客票务信息，诈骗分子再利用这些信息实施网络诈骗，骗取金额1000多万元。警方共抓获黑客林某等犯罪嫌疑人20名，缴获航空票务类公民信息30多万条和大量账号、密码信息。

同一黑客采用暴力破解等手段进行攻击



数据安全问题


已显示"民航旅客信息泄露"的搜索结果。 仍然搜索： 民航旅客信息泄漏
"最安全航空公司"出最大纰漏:940万乘客信息泄露,你在其...
2018年10月26日 能录入自身系统,必须得到当事人或者有关人员的确认,信息如何使用,乘客有权利知晓。"民航专家李小群对《每日经济新闻》记者表示,近千万级的航空乘客数...
315.hexun.com/2018-10-26/19500... 百度快照
保护旅客个人信息安全 民航责无旁贷-中国民航网
2017年3月30日 其次,要加强对旅客个人信息的管理监督。近年来,民航局高度关注旅客信息泄露事件,近年来,旅客信息保护水平也在不断提高,但仍有努力的空间。几个月前,国...
www.caacnews.com.cn/zk/zj/quny... 百度快照
泄露乘客信息,国航一员工被处理!_OpenLaw
2020年1月3日 1月3日中午有民航自媒体曝光,疑似中国国际航空公司一位员工在其个人微博晒出明星的乘机信息,包括出生日期、座位号、以及常旅客会员卡级别等情况。
搜狐网 百度快照
其他人还在搜
旅客信息事件 泄露旅客信息 航班信息泄露怎么办 中国民航信息 短信
怎样确定个人信息没有泄露 民航规定的重要旅客 航班信息总被泄露 南航旅客信息
山东:特大民航旅客乘机信息泄露案将被公诉-中国法院网
2015年4月28日 《法制日报》记者今天从济南市历城区人民检察院获悉,该院目前已对这起特大民航旅客乘机信息泄露案审查起诉完毕,并将于近期对高某等18名犯罪嫌疑人以涉嫌提供侵入计...
中国法院网 百度快照
特大民航旅客乘机信息泄露案将被公诉-中新网
2015年4月28日 《法制日报》记者今天从济南市历城区人民检察院获悉,该院目前已对这起特大民航旅客乘机信息泄露案审查起诉完毕,并将于近期对高某等18名犯罪嫌疑人以涉...


19-9-7 16:11 来自iPhone客户端
十关注
这个晚会配置还不错。
旅客姓名: 张杰
英文名: ZHANG
CHN
旅客姓名: 韩红
英文名: HANHC
09/26
CHN
旅客姓名: 周笔畅
英文名: ZHOUB
/07/26
CHN
旅客姓名: SHAOH
英文名: SHAOH
生日
旅客姓名: WENW
英文名: WENW
生日
南京 · 南京禄口国际机场

数据安全问题

提示　站外联系隐患：安全风险 | 请卖家注意:所有交易均设置付款后自动查看，内容为主题帖界面最后一个回复，以方便动态提示自己的付款买家 | 资金密码设置提醒 |

首页 ‹ 网站首页 ‹ 数据-情报类

225665条广州航空数据

主题帖交易信息一览

交易编号：	12213	商品单价：	0.0023 [BTC]	交易发布时间：	10-09 20:29	刷新本页	
交易类型：	出售	约合美金：	14.69 [美元]	投诉保护期限：	付款后 7 天	进入本主题公开评论区	
交易状态：	出售中	本单成交：	1	商家最后在线：	10-25 15:28		
出售数量：	60	剩余数量：	59	举报内容抄袭：	被抄袭的交易编号	提 交	

发布者信息一览[当前粗略统计]

账户名称：	12345012345	在售单数：	23	强制撤诉单数：	0	
账户编号：	36038	总出售额：	0.0062	强制退款单数：	0	请交易双方恪守诚信，该退款退款,该撤销投诉就撤销
注册时间	2018-08-29	总购买额：	0	强制退款总额：	0	如果网站强制撤销投诉或强制退款，将留下不太适当的信誉记录

您未设置资金密码，为资金安全，站内一切涉及资金动作，均需出示资金密码，请设置您的资金密码

	请确保你能记住的密码，首次设置,一日之内忘记密码可点击重置，逾期不负责.
密码要求	强烈建议用笔记下该密码，本站不受理密码丢失问题，因为无法证明你就是你！(要不您留个姓名电话?)
密码要求	大小写字母与数字任意组合,8位以上,20位以下，非字母与数字部分将被自动删除
资金密码	
再次输入	
	提 交
提示	不是登录密码，是付款购买,提币时用的资金密码
提示	非数字字母部分被清除，比如设置 asAS!@121212 将自动更改为 asAS121212

发起: 私密会话发起方:[12345012345].
参与: 私密会话参与方[qazwsx012345].

回复　1 帖子 • 分页： 1 / 1

12345012345
帖子： 124
注册时间： 2018年-8月-29日 18:06
联系：

225665条广州航空数据
由 12345012345 » 2018年-10月-09日 20:28

225665条广州航空数据
内容包含，电话号码，联系人，出发地，目的地，地址，邮编
数据并非最新，售出不以任何理由退款
自动发货

快速回复

尊敬的朋友：如下标题和内容是卖家对其商品的描述和承诺，请务必仔细阅读，防止错买，如有疑问，请在下面回复，卖家的补充回复同样视为交易承诺.

aa2116033
帖子： 23
注册时间： 2018年-11月-01日 05:56
联系：

全国旅客机票订票数据，每天一万条
由 aa2116033 » 2018年-12月-25日 14:04

1.出全国旅客订票数据，未起飞和历史数据都有 未起飞的数据每天大概更新一万条，日期和航空公司你定

2.内容包含：姓名，身份证，航班日期，起飞时间，到达时间，航空公司，航班号，始发站，目的站，电子票号，手机号

3.绝对真实数据，准确率100%

4.有疑问请点击呼叫卖家然后留言，卖家看到第一时间回复！！！

附件

	A	B	C	D	E	F	G	H	I	J	K
1	姓名	身份证	航班日期	起飞时间	到达时间	航空公司	航班号	始发站	目的站	电子票号	电话
2	吴山奇	610302199002091019	2018年12月25日	2018/12/25 7:00	2018/12/25 9:20	中国东方航空	航5137	上海--	北京	7818244367309	13818543240
3	王俊才	320982198710022293	2018年12月25日	2018/12/25 15:00	2018/12/25 18:00	中国东方航空	航5378	三亚--	上海	7818248915168	15761723776
4	何彬	310112198006111212	2018年12月25日	2018/12/25 20:00	2018/12/25 22:30	中国东方航空	航2168	上海--	西安	7818244382008	13611737595
5	张云	432801197202204022	2018年12月25日	2018/12/25 10:15	2018/12/25 11:20	中国东方航空	航5809	腾冲--	昆明	7818270279054	13827948166
6	曹京玉	110101196212092192	2018年12月25日	2018/12/25 6:50	2018/12/25 10:25	中国东方航空	航5181	北京--	广州	7818259010778	13651040289
7	沈毅辉	310110197802171251	2018年12月25日	2018/12/25 7:00	2018/12/25 9:20	中国东方航空	航5137	上海--	北京	7818244465149	13818431230
8	杨荣	532901199411292012	2018年12月25日	2018/12/25 14:10	2018/12/25 17:30	中国东方航空	航5455	宁波--	昆明	7818270918422	13624635586
9	苏朝旭	511527199604150026	2018年12月25日	2018/12/25 15:40	2018/12/25 19:15	中国东方航空	航5267	上海--	宜宾	7818271025673	15921886073
10	周维强	533023199705061017	2018年12月25日	2018/12/25 20:00	2018/12/25 21:15	中国东方航空	航5702	昆明--	腾冲	7818274496268	15025078237
11	周敦康	620102199907031114	2018年12月25日	2018/12/25 10:50	2018/12/25 13:05	中国东方航空	航2837	石家庄--	兰州	7818278304319	13519622759
12	曾巍	460102198703050638	2018年12月25日	2018/12/25 6:25	2018/12/25 9:15	中国东方航空	航5202	海口--	上海	7818278270207	18689774530
13	李志伟	340121199206025817	2018年12月25日	2018/12/25 21:00	2018/12/25 23:15	中国东方航空	航5436	成都--	合肥	7818274189874	18256505751
14	顾一磊	310105198003171638	2018年12月25日	2018/12/25 14:30	2018/12/25 16:55	中国东方航空	航5306	广州--	上海	7818244214654	13918532868
15	王博	650102199308274025	2018年12月25日	2018/12/25 15:10	2018/12/25 17:05	中国东方航空	航5739	昆明--	广州	7818274244823	18199810752
16	童昱	320205198908091228	2018年12月25日	2018/12/25 8:15	2018/12/25 9:35	中国东方航空	航5481	宁波--	青岛	7818274615590	18606697540
17	李君瑞	412829199012150045	2018年12月25日	2018/12/25 17:00	2018/12/25 19:00	中国东方航空	航5385	上海--	郑州	7818278365779	18530873623
18	李磊	610103198606261219	2018年12月25日	2018/12/25 21:05	2018/12/25 22:55	中国东方航空	航2886	西安--	南京	7818278262034	15991776009
19	罗九梅	460033199806134594	2018年12月25日	2018/12/25 7:15	2018/12/25 10:25	中国东方航空	航5620	哈尔滨--	上海	7818270524313	15561965021
20	何璐娟	142602198610030528	2018年12月25日	2018/12/25 13:10	2018/12/25 15:20	中国东方航空	航5192	运城--	上海	7818280067928	18635778774
21	刘呈婕	522424198103254824	2018年12月25日	2018/12/25 13:20	2018/12/25 15:45	中国东方航空	航9617	宁波--	郑州	7818280011910	18883989868
22	王宁	210181199309044022	2018年12月25日	2018/12/25 15:15	2018/12/25 18:30	中国东方航空	航2727	南京--	三亚	7818270495584	15004091259
23	李东锋	130421199202126015	2018年12月25日	2018/12/25 17:00	2018/12/25 19:20	中国东方航空	航0774	南宁--	南京	7818257897753	13652027578
24	徐俊豪	513422199606030014	2018年12月25日	2018/12/25 19:35	2018/12/25 22:10	中国东方航空	航5175	南昌--	北京	7818280092101	15579154994
25	陈学勇	320705194911222011	2018年12月25日	2018/12/25 11:00	2018/12/25 13:40	中国东方航空	航2870	广州--	遵安	7818280088230	13016908051

这是基本的数据格式
11111.png (59.71 KiB) 查看 754 次

相关法规标准

1、《网络安全法》
2、《个人信息保护法》(草案)
3、《数据安全管理办法(征求意见稿)》
4、《个人信息和重要数据出境安全评估办法(征求意见稿)》
5、《天津市数据安全管理办法(暂行)》
6、《个人信息安全规范》
7、《个人信息去标识化指南》(征求意见稿)
8、《个人信息安全影响评估指南》(征求意见稿)
9、《信息安全技术 数据安全能力成熟度模型》
10、《GDPR》

民航信息泄漏相关案件

原告通过去哪儿网购买了X航的机票后，收到与所购航班机票相关的疑似诈骗信息的短信，就此原告起诉北京趣拿信息技术有限公司（下称：去哪儿网）和X航空公司泄露其隐私信息，包括姓名、手机号及行程安排。

北京市第一中级人民法院于2017年3月27日作出（2017）京01民终509号民事判决：一、北京趣拿信息技术有限公司于本判决生效后十日内在其官方网站首页以公告形式向庞XX赔礼道歉，赔礼道歉公告的持续时间为连续三天；二、X航空公司于本判决生效后十日内在其官方网站首页以公告形式向庞XX赔礼道歉，赔礼道歉公告的持续时间为连续三天。

举证问题

作为原告来讲举证存在以下问题：

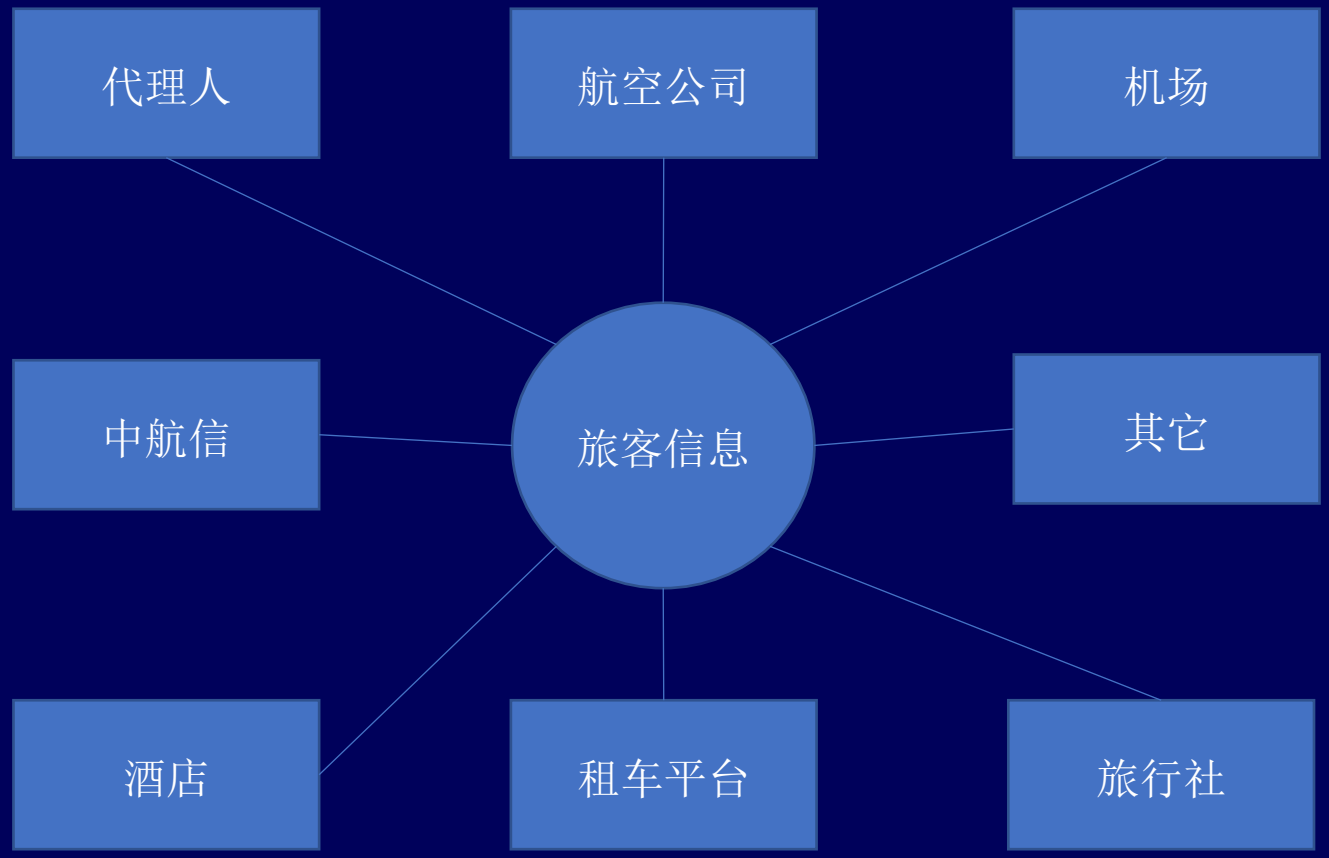
第一，证明信息泄露可能需要具备一定的技术知识或条件，而作为个人的原告通常不具备。

第二，原、被告实质上存在地位不对等。个人信息泄露引发的侵权纠纷案件中，原告均为自然人，而被告通常是具有优势地位的企业。在这样的不对等的关系下，原告的取证受到了极大的限制。

第三，现行法下，个人无从了解自身个人信息收集、使用和保护状况。

在庞XX案中，法院认为“客观上，法律不能也不应要求庞XX确凿地证明必定是东航或趣拿公司泄露了其隐私信息。”

旅客信息可能泄漏的渠道



涉事单位如何“自证清白”

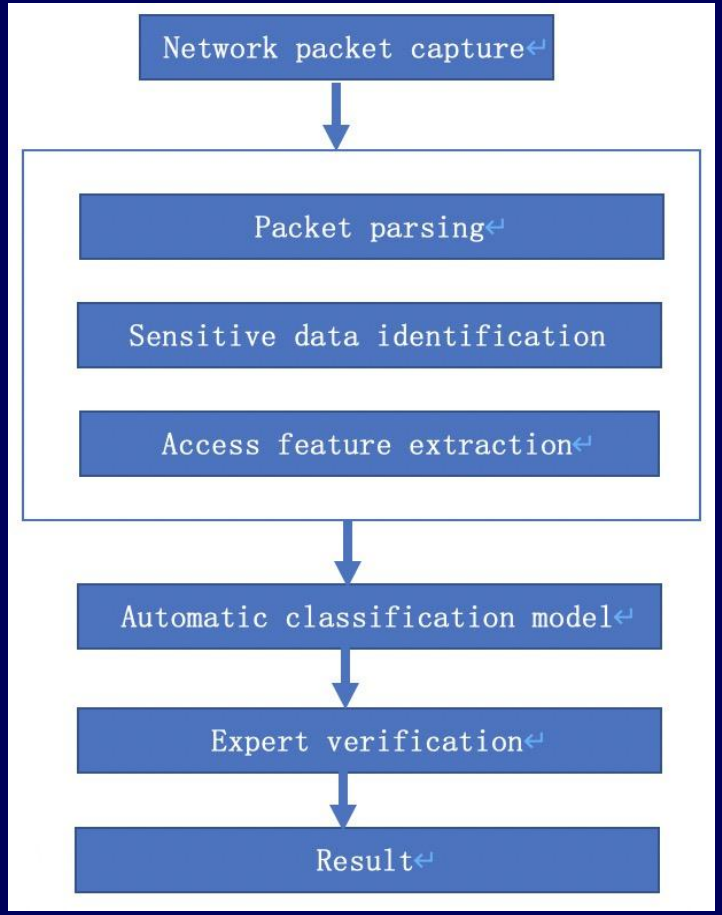
- 在上述案件中，相关企业也没有拿出完全证明自己清白的证据，那遇到类似事件应该怎么处理或者未雨绸缪呢？

- 除了采用数据加密等保护措施外，数据安全审计也是一种非常必要的手段，不但是加强企业自身的数据安全管理水平的手段，也是相关监管部门的要求。

- 那么对于一个高度依赖信息化的企业，信息系统数据众多、应用系统格式不统一，如何集中对应用系统日志进行审计分析？结合民航实际情况提出通过对网络流量进行分析、处理，实现对多个应用系统敏感数据的使用的集中审计，提高民航各单位的数据安全管理能力。

总体架构

通过对网络流量进行采集、解析，然后对网络流量中敏感信息进行识别，并采用机器学习的方法对敏感信息的访问行为进行分类，识别针对敏感数据的恶意访问行为。



民航旅客信息字段

对数据包内容进行解析，识别网络数据包中是否含有敏感数据，以及按照敏感数据分类模型中的方法进行敏感数据的分级。

对于民航旅客信息可能包括：姓名、身份证、家庭住址、电话号码、微信账号、Email账号、银行卡号、航班号、电子客票号码、航班时间、出发地、目的地等内容。

敏感数据分级分类

- 民航旅客分级分类方法（研究中）

（1）非常敏感信息

身份证、电话号码、姓名、电子客票号码 不法分子获取上述信息后，不但可以实施机票诈骗，还可能对 旅客造成其它方面的困扰。（包括未满十四周岁的未成年人）

（2）比较敏感信息

家庭住址、银行卡号、即时通信工具账号、电子邮件账号 不法分子如果仅利用这些信息可用于人肉搜索等。

（3）一般信息

航班号、航班时间、出发地、目的地等。 不法分子如果仅利用这些信息仅可用于旅客的数量统计等。

民航旅客信息识别方法

（1）使用模式匹配方法，匹配数据的长度、字符类型和格式；

（2）使用关键字匹配，对结构化数据的标头进行识别；

（3）使用关键字+模式匹配的方式进行识别，通过识别文件中的关键字例如：邮箱等关键字，然后在关键字附件进行模式匹配，识别是否存在敏感信息。

（4）采用NLP识别方法，识别数据包中的姓名或地址信息等。

自动分类方法

采用机器学习算法对网络中针对敏感数据的访问行为进行分类。通过对网络中敏感数据访问的时间、敏感数据的数量、频率、IP地址等特征进行分析，采用机器学习算法对这些数据进行自动分类。

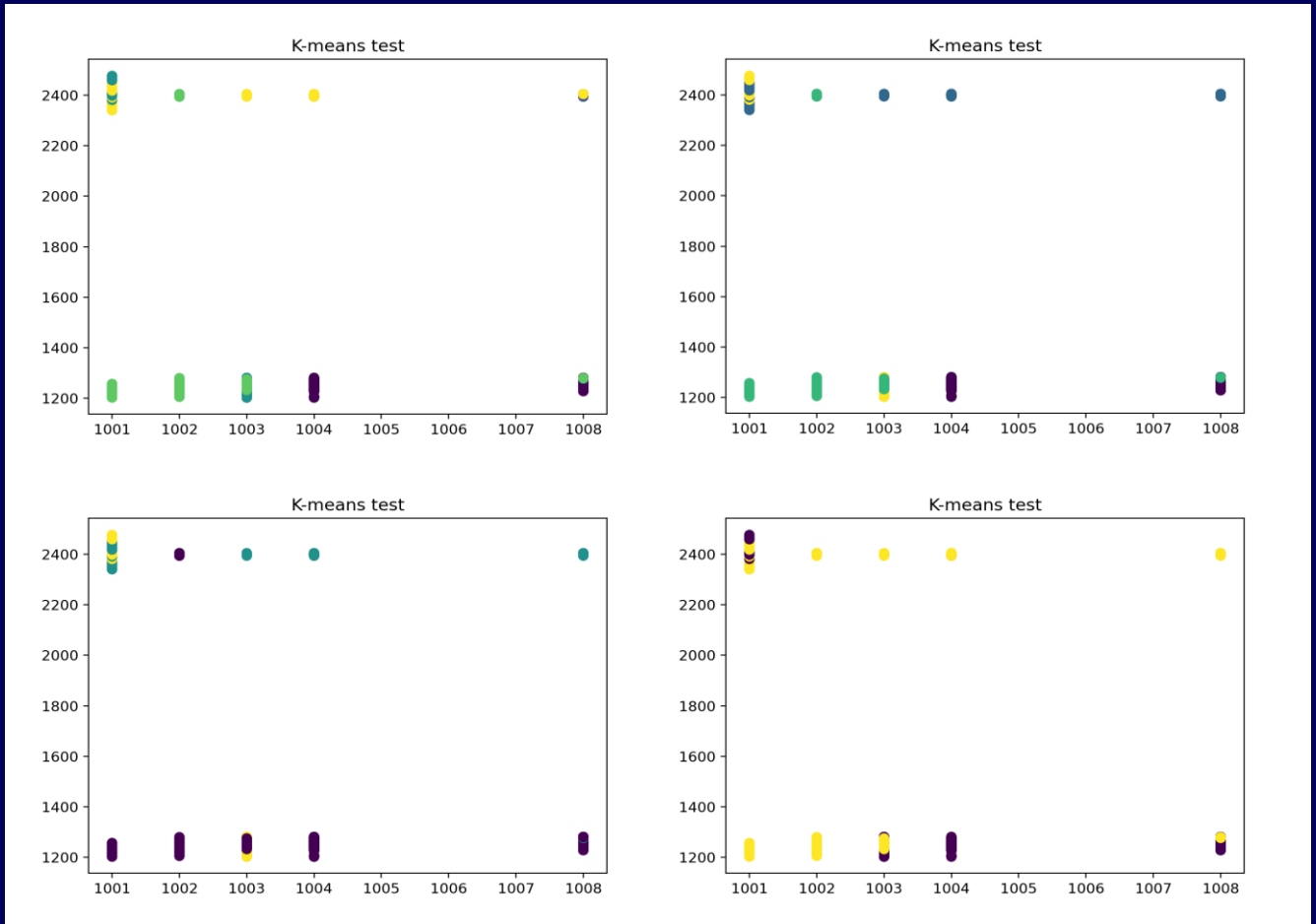
1、数据预处理

本次数据分析只针对数据访问的源地址、目的地址、访问时间、访问频率、访问数据量、访问用户的权限进行分析。

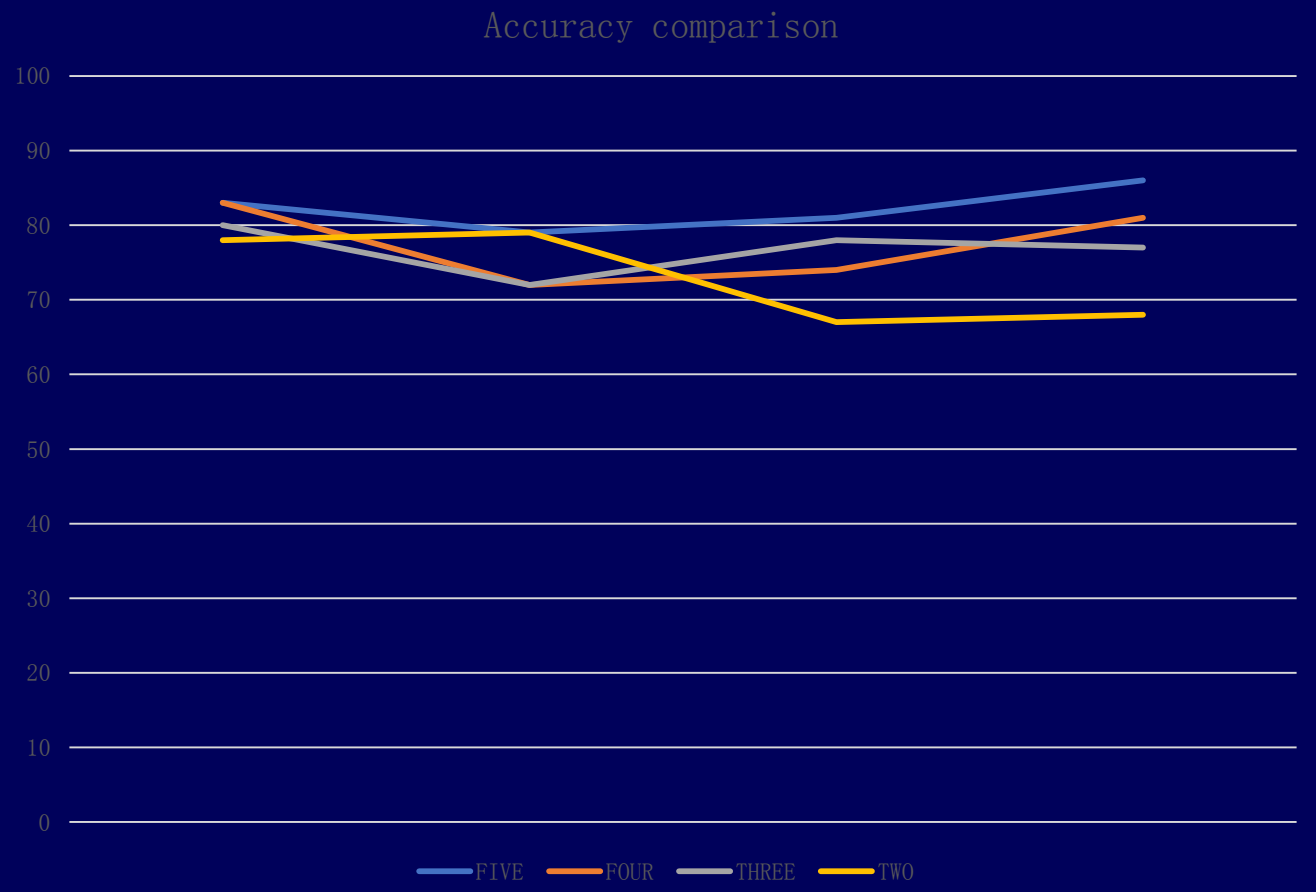
2、数据分析

采用K-means算法对网络中针对敏感数据的访问行为进行分类。通过对网络中敏感数据访问的时间、敏感数据的数量、频率、IP地址等特征进行分析，采用机器学习算法对这些数据进行自动分类。

K-means识别算法


K-means test
2400
2200
2000
1800
1600
1400
1200
1001
1002
1003
1004
1005
1006
1007
1008
K-means test
2400
2200
2000
1800
1600
1400
1200
1001
1002
1003
1004
1005
1006
1007
1008
K-means test
2400
2200
2000
1800
1600
1400
1200
1001
1002
1003
1004
1005
1006
1007
1008
K-means test
2400
2200
2000
1800
1600
1400
1200
1001
1002
1003
1004
1005
1006
1007
1008

K-means识别算法


Accuracy comparison
100
90
80
70
60
50
40
30
20
10
0
FIVE
FOUR
THREE
TWO

自动分类结果

采用K-Means划分的五类，如下表：

数据源地址	数据目的地址	访问数据量	访问频率	用户访问权限
1008	2211.83	15833.33	91.83	1333.33
1006.19	1252.88	5172.84	832.35	1039.51
1002.28	2397.85	11705.92	265.39	1065.79
1002.45	1608.23	766.67	789.43	1000
1002.49	1608.22	7834.75	736.81	1136.75

自动分类结果

对分类数据进行分析，以下为异常类

数据源地址	数据目的地址	访问数据量	访问频率	用户访问权限
1008	2211.83	15833.33	91.83	1333.33
1002.28	2397.85	11705.92	265.39	1065.79

即：上述为圆心的数据子集为异常访问

长风破浪会有
时，
直挂云帆济沧
海。

网络安全创新大会
Cyber Security Innovation Summit

THANKS