阿图因软件空间测绘系统的业务安全实践

演讲者：王连赢

腾讯安全玄武实验室简介

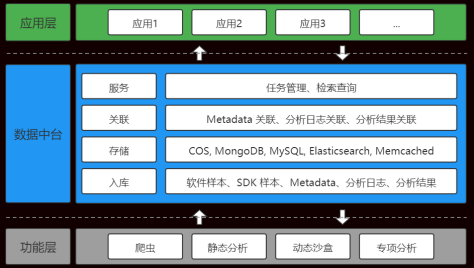
- 成立于 2014 年
- 外部安全赛事，内部红蓝对抗
- 对外前沿安全研究，对内业务安全支持
- 三十余项研究成果在国际安全会议上发表
- 对外报告漏洞上千个
- 多次发现影响全行业的通用安全问题

阿图因项目简介

- 软件空间安全测绘
  - 网络空间测绘：Shodan, ZoomEye, FOFA
  - 软件空间测绘：阿图因
- 跨平台
- 全流程自动化
- 千万级软件样本量
- 对外影响力
  - 2017 世界互联网大会 58 项世界领先科技成果
  - 多次 CNCERT 联合应急响应

阿图因系统概览

- 整体设计
  - 围绕数据中台
  - 统一数据格式
  - 通过沙盒消除平台差异
- 数据规模
  - 各类样本 5000 万+
  - 各类原始文件存储 700 TB+
  - 各类分析结果数据 1 TB+
- 工程实现
  - 云原生



安全研究与业务落地的相互促进

1. 供应链安全

阿图因典型案例回顾

供应链安全 – 概述

- 背景
  - 软件集成度越来越高
  - 供应链依赖越来越强
- 供应链安全大事纪
  - 2015.09 XcodeGhost
  - 2017.08 Xshell 后门
  - 2017.09 CCleaner 后门
  - 2018.03 竹节虫后门
  - 2018.12 驱动人生升级服务攻击
  - 2019.03 ShadowHammer 华硕升级服务攻击
  - 2020.12 SolarWinds Orion 软件后门

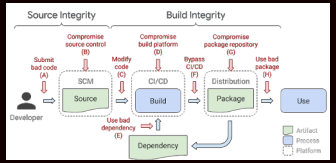


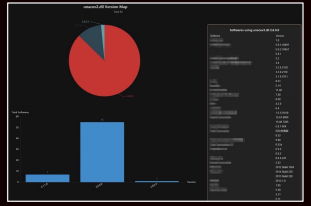
图. Google SLSA 框架供应链威胁描述

图. 供应链安全风险概览

供应链安全 – 案例 1：WinRAR RCE 漏洞

- 发现者
  - Check Point 公司安全研究员
- 问题模块
  - UNACEV2.DLL
- 问题类型
  - 逻辑漏洞：路径穿越

- 阿图因影响范围评估



供应链安全 – 案例 2：WhatsApp RCE 漏洞

- 发现者
  - 新加坡安全研究员 Awakened
- 问题模块
  - android-gif-drawable
- 问题类型
  - 内存漏洞：UAF

- 阿图因影响范围评估
  - 阿图因自动化反查
    - 可能受影响的 APK 10 万+
    - 可能受影响的 App 4 万+

供应链安全 – 案例 3：BucketShock 云存储漏洞

- 问题描述
  - 云存储应用越权访问和文件上传漏洞
- 发现者
  - 玄武实验室研究员
- 问题模块
  - 众多云存储 SDK
- 问题类型
  - 逻辑漏洞

- 阿图因影响范围评估
  - 阿图因自动化反查
    - 国内使用主流云存储服务的 Android 应用 4 千余个
  - 人工抽样验证
    - 上述应用中约 70% 受影响

供应链安全 – 案例 4：Xshell, CCleaner 事件

- 攻击方式
  - 攻击控制研发编译环境
  - 插入恶意代码动态加载 Shellcode
  - Shellcode 内实现攻击
- 阿图因检测方案及结果
  - 未知恶意行为判定
  - 可以不借助任何先验特征，准确检测出 XshellGhost, CCleanerBackdoor 问题

供应链安全 – 案例 5：驱动人生恶意代码事件

- 攻击方式
  - 厂商升级服务器被黑
  - 云控下发恶意代码
- 受影响的更新服务地址
  - dtlupdate.updrv.com
  - globalupdate.updrv.com
  - dtl.update.updrv.com

- 阿图因影响范围评估
  - 自动反查访问过该域名的软件
  - 解决网络行为与软件主体对应关系问题

2. 数据隐私安全

阿图因数据隐私合规腾讯内部实践

数据隐私安全 – 背景

- 行业现状混乱无序，监管压力日趋增强
- 国内数据安全监管大事纪
  - 2019.01 四部委发布《关于开展APP违法违规收集使用个人信息专项治理的公告》
  - 2019.03 APP专项治理工作组发布《APP违法违规收集使用个人信息自评估指南》
  - 2019.05 国家网信办发布《个人信息出境安全评估办法（征求意见稿）》
  - 2019.08 国家网信办发布《儿童个人信息网络保护规定》
  - 2019.12 四部委联合发布《APP违法违规收集使用个人信息行为认定方法》
  - 2020.08 工信部纵深推进APP侵害用户权益专项整治行动
  - 2020.08 四部委启动2020年APP违法违规收集使用个人信息治理
  - 2021.06 十三届全国人大常委会第二十九次会议通过了《数据安全法》
  - 2021.08 十三届全国人大常委会第三十次会议通过了《个人信息保护法》

中华人民共和国国家互联网信息办公室
Cyberspace Administration of China

中华人民共和国工业和信息化部
Ministry of Industry and Information Technology of the People's Republic of China

全国人民代表大会
The National People's Congress of the People's Republic of China

数据隐私安全 – 技术拆解

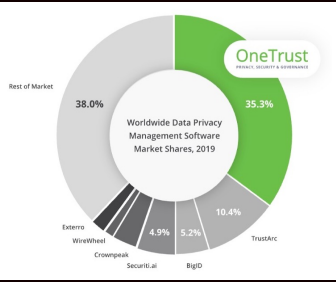
- 隐私数据识别
- 程序分析（静态、动态）
- SDK 检测
- SDK 行为分析

- 传输主体身份校验
- 数据加密
- 传输信道安全
- 传输接口安全

- 数据的加密存储
- 算法、密钥管理
- 数据的可用性、完整性

- 数据脱敏
- 使用授权管理
- 数据流转监控
- 数据防泄露
- 数据溯源

数据隐私安全 – 行业现状



爱加密 www.ijiami.cn

棚棚安全 BANGCLE

指掌易

百度智能云 cloud.baidu.com

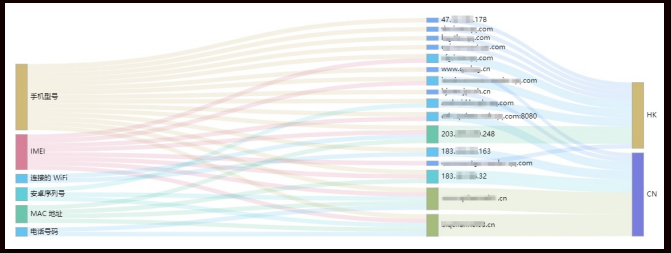
奇安信 新一代网络安全领军者

全知科技 QUAN ZHI TECH

数据隐私安全 – 阿图因核心能力概览

数据隐私安全 – 端上的隐私数据全生命周期流动跟踪

- 隐私数据从什么位置读取?
  - 调用栈
- 中间经过了什么操作?
  - 加密、压缩等操作
- 发送到了什么地址?
  - 域名、IP 地址
- 数据是否出境?
  - 国家区域



数据隐私安全 – 大规模高覆盖度 SDK 检测

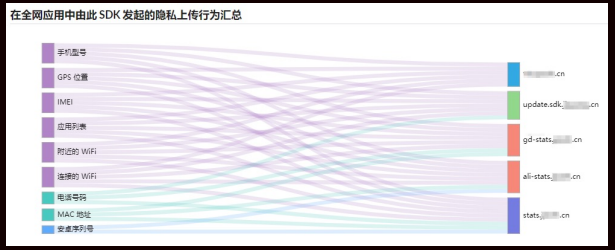
检测方案	核心方法	优点	缺点	适用场景
基于聚类	大规模聚类	1. 无需 SDK 样本 2. 漏报率低 3. 大规模、自动化	无法可靠获取版本信息	自主分发渠道、私有 SDK
基于样本	代码特学习、构造、匹配	1. 可以准确检测版本 2. metadata 标准化	1. 需要预先构建样本库 2. 无法检测未知 SDK 3. 样本库需要持续维护	标准分发渠道

千万级 SDK 样本集

Android / iOS

数据隐私安全 – SDK 数据隐私治理

- SDK 隐私行为分析
  - 如何解决行为触发覆盖率问题?
  - 如何区分 App 行为与 SDK 行为?
- SDK 隐私行为全局视角
  - 不同开发者使用方式不同
  - 不同环境表现不同



数据隐私安全 – 端到端的解决方案

- 技术层面的合规并不是终点
- 需要解决服务对象的实际问题
- 不同产品对象业务目标不同
- 不同业务目标对应不同的合规需求
  - App 产品的业务目标和合规需求是什么?
  - SDK 产品一样吗?

数据隐私安全 – 案例 1：SDK 资产盘点

- 业务场景
  - SDK 资产盘点
- 业务价值
  - 对于 App 产品部门
    - App 接入了哪些 SDK?
    - 使用了什么版本？是否最新版?
  - 对于 SDK 产品部门
    - SDK 被哪些 App 使用?
    - 版本分布情况如何?
  - 对于数据隐私合规
    - 无用组件清理，收敛数据隐私风险

“xxx SDK 目前产品已不需要使用，计划在1月8日左右完成删除”

“这三个我们后面准备删了，现阶段意义不大”

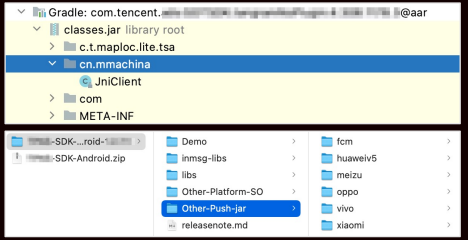
“过完年马上就删，三月份的版本就看不到了”

“xxx 看下需要排期删了 xx、xx、xx 相关的无用代码哈”

“我来了4年了，没听说过这个东西，我问问离职的同事吧”

数据隐私安全 – 案例 2：SDK 嵌套问题

- SDK 集成嵌套其它 SDK
  - 常用于功能整合
  - 常见于广告类、推送类、性能优化类 SDK
- 解决方案
  - 保持 SDK 特征数据的原子性
  - 复合 SDK 特征清洗掉原子特征
  - 检测结果给出原子数据，保证准确性
  - 自动学习构建 SDK 嵌套关系图，解决组件引入溯源问题



数据隐私安全 – 案例 3：SDK 正常功能被恶意使用

- 腾讯某异常上报 SDK 背锅案例
  - 该 SDK 具有热更新功能
  - App 使用热更新动态加载代码
  - 动态加载的代码中读取了用户隐私数据
  - 隐私数据通过回调接口上传到了腾讯服务器
- SDK 产品团队回应与处理方案
  - 确认了问题与设计层面原因
  - 修改隐私声明，澄清责任界面
- FYI
  - 热更新方便了开发者的同时也带了诸多风险
  - Google Play / Apple App Store 明确禁止热更新

```
at android.app.ApplicationPackageManager.getInstalledPackages(Native Method)
    at xx.xx.xx.android.utils.DeviceInfo.c(Unknown Source)
    at xx.xx.xx.android.utils.DeviceInfo.init(Unknown Source)
    at xx.xx.xx.android.api.JShareInterface.init(Unknown Source)
    at xx.xx.xx.toothbrush.XrushApp.onCreate(XrushApp.java:103)
    at com.tencent.xx.beta.xx.XxxPatchReflectApplication.onCreate(BUGLY:189)
    at com.wrapper.proxyapplication.WrapperProxyApplication.Ooo0ooO0oO(d)
    at com.wrapper.proxyapplication.WrapperProxyApplication.onCreate()
    at xx.xx.xx.toothbrush.MyWrapperProxyApplication.onCreate()
    at android.app.Instrumentation.callApplicationOnCreate()
```

“我们提供了 error、崩溃回调接口，用户可以放入自己采集的信息一起传到我们的后台” —— 该 SDK 产品团队确认问题原因


（一）App开发者和/或终端用户主动提供或我们直接收集的个人信息

为实现各类[redacted] SDK的特定业务功能，我们可能需要向App开发者和/或终端用户收集相关个人信息，主要包括：日志信息（包括：第三方开发者自定义日志、Logcat日志以及APP崩溃堆栈信息）、设备ID（包括：androidid以及idfv）、联网信息、系统名称、系统版本以及国家码。

3. 仿冒 App 打击

阿图因大规模数据分析方法应用

仿冒 App 打击 – 概述

- 背景
  - 国内 Android 应用市场管理混乱
  - 应用分发、传播渠道众多
  - Android 系统开放，重打包技术门槛低
- 应用场景
  - 打击盗版
  - 打击游戏私服
  - 打击恶意应用
  - 打击新型网络犯罪
    - 金融诈骗、杀猪盘
    - 裸聊勒索
  - …

仿冒 App 打击 – 行业实践

- 通过管理手段解决问题
  - 移动应用备案机制
  - 金融、教育等强监管领域
- 相关文件
  - 中国人民银行关于发布金融行业标准加强移动金融客户端应用软件安全管理的通知（银发〔2019〕237号）
  - 教育部办公厅关于印发《教育移动互联网应用程序备案管理办法》的通知（教技厅〔2019〕3号）

中国人民银行文件

银发〔2019〕237号

中国人民银行关于发布金融行业标准
加强移动金融客户端应用软件安全管理的通知

中国人民银行上海总部，各分行，营业管理部，各省会（首府）城市中心支行，各副省级城市中心支行；国家开发银行，各政策性银行、国有商业银行、股份制商业银行，中国邮政储蓄银行；各证券公司，基金公司、期货公司、私募股权基金管理机构；各保险集团（控股）公司、保险公司、保险资产管理公司；中国银

移动金融客户端应用软件备案专栏

首页>>>>正文：App备案专栏 >通知公告

移动金融客户端应用软件实名备案名单公布（第一批）

2020年06月03日

根据《中国人民银行关于发布金融行业标准加强移动金融客户端应用软件安全管理的通知》（银发〔2019〕237号），中国互联网金融协会（以下简称协会）承担移动金融应用客户端软件（以下简称客户端软件）实名备案工作，经综合审核并组合相关机构测评和公示情况，形成了第一批完成客户端软件实名备案名单(共73款)，现予以公布。

中国互联网金融协会

2020年6月3日

序号	App图标	App产品名称	责任名称	产品类型	版本号
1		中国工商银行	中国工商银行股份有限公司	Android	5.1.8.1.0
2				IOS	5.1.8.1.0
3		民生银行	中国民生银行股份有限公司	Android	5.21
4				IOS	5.21
5		民生信用卡	中国民生银行股份有限公司	Android	6.2
6				IOS	6.3

仿冒 App 打击 – 阿图因实践

- 通过技术手段解决问题
  - 大规模、自动化
  - 解决长尾应用问题
- 核心方法
  - 大规模聚类
  - 大规模应用行为相似性分析

3800万+

阿图因 Android 应用样本库规模

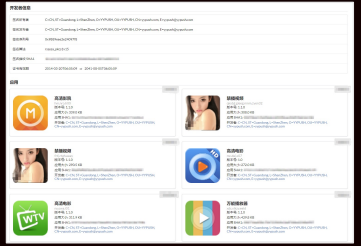
120万+

阿图因 iOS 应用样本库规模

仿冒 App 打击 – 案例

• 盗版应用统计

• 盗版开发者反查



仿冒 App 打击 – 案例

- 开发者真实信息溯源
  - 以服务平台 SDK 为切入点
  - 反向数据流跟踪自动提取开发者 ID
  - 自动关联服务平台工商信息
  - 自动关联相关应用
  - 通过服务平台调查取证
- 打破网络虚拟世界和现实世界的次元壁

总结

- 核心能力建设
- 有数据，有技术
- 懂安全，懂业务
- 方法论
  - 通过持续测绘将软件安全问题转化为数据查询
  - 通过数据分析方法弥补程序分析的不足
  - 通过数据的多维度关联发现未知问题

安全能力

反编译、逆向、脱壳、通用 Hook

数据流分析、 IR Lifting

…

数据能力

数据关联、数据聚类、特征学习

…

研发能力

虚拟化、分析沙盒、流式计算

…

感谢观看！

KCon 汇聚黑客的智慧

知道创宇 | KCon