APK解析

1.APK文件解压缩:

使用mimizip库, 不将文件解压缩到文件系统, 流式解析方法, 解析具体文件时直接操作内存. 应用签名存在两个版本, V1版本需要解压缩APK获取签名/校验数据等信息, V2直接在APK文件的zip编码区中解析.

2.*.RSA文件解析

从*.RSA文件中解写出应用开发个人/厂商的公钥签名与加密的校验文件hash数据. *.RSA文件为PKCS7标准二进制DER格式编码, 使用openssl crypto库解析. 数据校验/解密算法也使用crypto库.

3.AndroidManifest.xml解析解析

此文件并不是xml文本, 文件格式为AXML二进制编码.需要先将其解码为文本xml文件再解析内容.解码工作需要编码完成. 文件中包含了Android api版本, 应用权限等新信息. xml解析库使用pugixml.

4.resources.arsc文件解析

资源文件.包含应用中使用的字符串等信息.类似AXML文件,需要编码完成完成解析.

5.json文件解析

使用rapidjosn库

6.dex文件

虚拟机可执行文件. 需要解析获取应用的API列表.

7.其他

应用完整性检查

通过APK的校验机制进行完整性检查, 防止APK文件被篡改.无法通过完整性检查则报异常可疑APP.

应用签名白名单

建立应用签名白名单列表. AMC引擎检测样本为白名单则不进行后续检查.

关键字扫描

扫描APK包中的文本文件, 发现色情/暴力/反动词汇认为APK异常.通过辨别算法识别文本文件是否为应用本身的关键字词库. 需准备词库.

基于应用API与应用权限等信息识别

使用机器学习方法, 通过大量样别的API与应用权限等不同维度信息, 训练算法模型. AMC引擎通过训练的模型推断出APK文件是否为恶意APP. 待研究.

其他

AMC恶意程序引擎通过静态库的形式提供使用, 同时开发AMC_scanner等程序辅助线下样本分析.