

# 软件综合实践第一周进度报告

|        |    |   |    |     |  |
|--------|----|---|----|-----|--|
| 项目名称   |    | 基于深度学习和大模型的 android 恶意软件检测系统  |    |     |  |
| 项目成员   | 学号 | 20221982  | 姓名 | 潘钊葳 |  |
| 组内其他成员 | 学号 | 姓名  | 学号 | 姓名  |  |
|        |    |   |    |     |  |
|        |    |   |    |     |  |
|        |    | <p>1. 对 apk 进行解析：首先收集到了近六百个 apk 文件，和其对应的分类标签，涵盖了 gamble、scam、sex、white、black 五个类别，利用 androguard 对 apk 进行了解析，得到了每个 apk 的如下特征，并输出成 csv 文件：ID,App Name,Package Name,Main Activity,Activities,Services,Receivers,Permissions,Signing Scheme,Issuer,Subject,Serial Number,Hash Algorithm,Signature Algorithm,SHA1,SHA256,MD5,Call Graph Path</p> <pre> ID,App Name,Package Name,Main Activity,Activities,Services,Receivers,Permissions,Signing Scheme,Issuer,Subject,Serial Number,Hash Algorithm,Signature Algorithm,SHA1,SHA256,MD5,Call Graph Path aditt_android_0828.apk,爱情岛论坛,com.youzi.bigurl,com.youzi.bigurl.MainActivity,[[]],[[{"android.permission.INTERNET","android.permission.WRITE_EXTERNAL_STORAGE"}],both,Common Name: xiaoxiong,Organizational Unit: zhongxing,Organization: zhongxing,Locality: chengdu,State/Province: sichuan,Country: 610000,"Common Name: xiaoxiong,Organizational Unit: zhongxing,Organization: zhongxing,Locality: chengdu,State/Province: sichuan,Country: 610000",sha256,rsa256_pkcs1v15,b"\xd7f1x9d\xbd\x97\xbd\xbd\xba\xfd\x96\xba\xbd\xbc?*\e\x81","b"\xfcf\xba\xea5\x22\xfc\x0371xdc.\xf87j\k7f\x9b\x92\webe.\kcf\xcf\xdc4c\kaf\vef\xb7q\x5\x2a\xfi\x8\xcd",9e94e483ddc324d564a998bb72bdf,backdataset\white\call_graph_aditt_android_0828.apk.txt </pre> <p>2. 构建数据集：根据得到的分类标签表，给每条数据打上了 result 标签，构造了一个含有 519 条数据的数据集：</p> <pre> ID,App Name,Package Name,Main Activity,Activities,Services,Receivers,Permissions,Signing Scheme,Issuer,Subject,Serial Number,Hash Algorithm,Signature Algorithm,SHA1,SHA256,MD5,Call Graph Path,result 15060399.apk,少年客,com.chuckergames.kungfuguaji.org.cocos2dx.cpp.AppActivity,[{"org.cocos2dx.cpp.AppActivity","com.chuckergames.kungfuguaji.wxapi.WXPayEntryActivity","org.cocos2dx.cpp.PayActivity","com.alipay.sdk.app.HSPayActivity","com.alipay.sdk.app.HSAuthActivity","com.alipay.sdk.app.PayResultActivity","com.alipay.sdk.app.AlipayResultActivity"}],[],[{"android.permission.ACCESS_WIFI_STATE","android.permission.READ_PHONE_STATE","android.permission.WAKE_LOCK","android.permission.ACCESS_NETWORK_STATE","android.permission.INTERNET"}],both,"Common Name: Shi Hu,Organization: Chucker Games","Common Name: Shi Hu,Organization: Chucker Games",0x340c149f,sha256,rsa256_pkcs1v15,b"\(xd8\x2cx885\xd7\x1f\x22\xeb\x0e\xae\x95\xfbz)\xd7\x96\xfe","b"\xfcf\xae\x72k\l9g\x5\xff\x7\xbd\x94\x95\x05\x94\x0f2\x0c\xbd\x81\x82b"?\xf7e\x9f\x9d\x8",0cf4e6841d2ab6f3dc67e2b1914f856,backdataset\batch1\call_graph_15060399.apk.txt,black </pre> <p>3. 设计了调用大模型进行 apk 特征分析的 prompt 模板，如下是使用该模板后大模型对某个 apk 特征的返回结果：</p> <pre> 对于给定的APK特征，按照您提供的模板来分析：  ### （一）APK 基本信息 1. 应用名称: 小鸭子    包名: com.yu.game.floatmenu_demoYz    主活动: com.stub.stub01.StartActivity 2. 活动列表    - com.yu.game.floatmenu_demoYz.MainActivity    - com.stub.stub01.StartActivity    - com.ubix.ssp.open.com.UBIXVideoActivity    - com.ubix.ssp.open.com.UBIXWebViewActivity    - com.ubix.ssp.open.com.AdActivity    - com.beizi.ad.AdActivity    - com.qihoo.ak.landingpage.LandingPage    - com.qihoo.ak.click.diaolog.DownloadDialogActivity 3. 服务列表    - com.ubix.ssp.open.com.DownloadService    - com.beizi.ad.DownloadService    - com.yu.game.floatmenu_demoYz.MyAccessibilityService 4. 接收器列表    无 </pre> |    |     |  |

```
5. 权限列表
- android.permission.READ_EXTERNAL_STORAGE
- com.android.launcher.permission.INSTALL_SHORTCUT
- android.permission.WRITE_EXTERNAL_STORAGE
- android.permission.SYSTEM_ALERT_WINDOW
- com.android.launcher.permission.UNINSTALL_SHORTCUT
- android.permission.WRITE_SETTINGS
- com.android.launcher.permission.WRITE_SETTINGS
- com.android.launcher.permission.READ_SETTINGS
- android.permission.INTERNET
- com.yu.game.floatmenu.demokyz.permission.BEIZI_AD

6. 签名信息
- 签名名: Xyz
- Common Name: Xyz
- Organizational Unit: Xyz
- Organization: Xyz
- Locality: Xyz
- State/province: Xyz
- Country: Xyz

7. 证书哈希
- 哈希算法: sha256
- 哈希值: 0x1
- 证书指纹: b'~Z\\xdd\\x9b\\xc6\\xd3\\xad\\x72\\x17\\xdc\\xd5\\x8e\\xd2f9\\xa18'
```

### (二) 综合分析

1. 应用主要功能

该应用可能是一个带有浮动菜单的游戏应用，包含广告显示、下载服务和系统设置修改等功能。

2. 活动和服务分析

应用包含多个活动和服务，主要涉及游戏主界面、广告展示和下载管理等功能。

3. 权限风险

应用请求的权限较多，包括读写存储、系统级悬浮窗口和网络访问等，存在滥用风险。

4. 签名和证书分析

应用由一个名为Xyz的证书所签名，但具体信息不明，安全性和可信度有待进一步评估。

5. 安全评估建议

- 总体安全评估：中等风险
- 建议用户在安装前审查应用权限，并注意不信任的行为，如显示悬浮窗口和未知的网络活动。

6. 建议

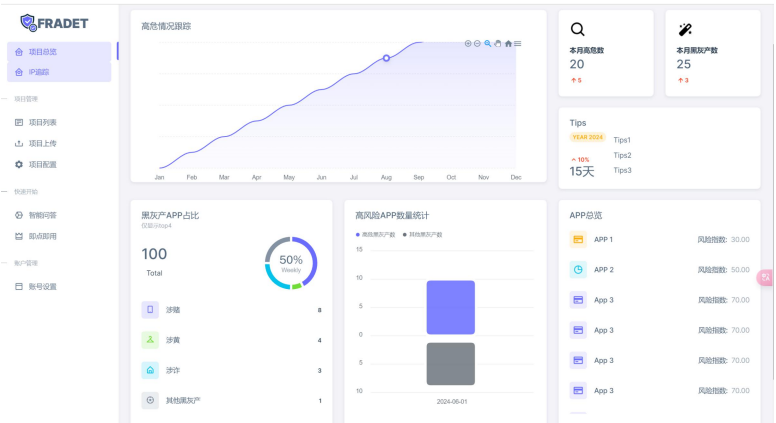
- 用户在安装前请谨慎评估应用的安全性。
- 避免在不信任的网络环境下使用该应用。
- 定期检查和管理应用权限，以确保安全使用。

PS C:\Users\86152\Desktop>ls -la

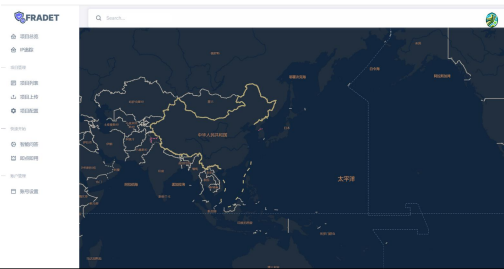
4. 构建了黑白名单数据库，并设计了接口供前端调用：

| SELECT * FROM apk LIMIT 100 |                   |              |                                   |       |  |
|-----------------------------|-------------------|--------------|-----------------------------------|-------|--|
| ID                          | packageName       | apkName      | md5                               |       |  |
| int                         | varchar(255)      | varchar(255) | varchar(255)                      |       |  |
| 1                           | ant0531           | 蚂蚁加速器        | 912d306c6fe3efb57285505af9f2af654 | black |  |
| 2                           | 525027            | 牡丹直播         | e09e4791ff323af6bf688d6c6b29b918  | black |  |
| 3                           | 712020            | 青青果视频        | 79a94a3896f0fdbf99765daffa07a02   | black |  |
| 4                           | 630021            | 青青直播         | c7a933628dad08f87411455f79998356  | black |  |
| 5                           | 712018            | 香肠视频         | ee6ed6edecd9669ced07f6834fe2dadd  | black |  |
| 6                           | 10                | 小创游戏库        | e55e8a3ad9f4a94addf985a71ae99175  | black |  |
| 7                           | 713011            | 夜月直播         | baff39ad0214ca46ae8197a107ea5bf2  | black |  |
| 8                           | 7539a28f18fb54af5 | 正宗南通长牌       | f1325c19af839335d87296d684335415  | black |  |
| 9                           | qqhh              | QQMM         | de5ef92d0437d5c77fd052b5602441b6  | black |  |
| 10                          | qqww              | QQMM         | de5ef92d0437d5c77fd052b5602441b6  | black |  |
| 11                          | tvchong2.3.6      | 虫虫影视         | 90af9915a9891fb1d232997a7926ddcf  | black |  |
| 12                          | WindmillVPN_v2.61 | WindmillVPN  | 64349d26a98f1c6a672a38e2e635bd7a  | black |  |
| 13                          | www               | 欣然业务网        | 47d8959037f897f1d854fda367f1eacf  | black |  |
| 14                          | xctvtbb969305     | 小曾TV台标版      | 1c610fbf12adebee0fa366a1edac903b  | black |  |
| 15                          | yxdmjsq0004       | 旋风加速器APP     | 62b50c8569173c72a574b1fdef1f2ed16 | black |  |

5. 修改前端页面，添加有关数据分析相关的内容，包括了高危情况跟踪、黑灰产占比、本月高危数、本月黑灰产数、大模型分析后提供的 tips、风险 app 总览：



6. 提供了对黑灰产业链中涉及的 IP 地址进行跟踪和可视化展示的功能，帮助用户更直观地了解和分析这些非法活动的网络结构和关联：



(1) 项目列表：显示已检测的 APK 列表，包括项目名称、检测时间、漏洞数量、修复进度和状态等信息，用户可以一目了然地了解各项目的当前状态和进展情况。

(2) 项目上传：提供了多种上传方式，包括通过链接、二维码和本地上传，方便用户根据不同需求选择合适的上传方式进行 APK 文件的上传。

| 模型介绍          |                        |                        |           |                   |                                     |
|---------------|------------------------|------------------------|-----------|-------------------|-------------------------------------|
| 模型            | 精度                     | 速度                     | 推荐        | 类型                | 可部署                                 |
| ReGVD         | <div><div></div></div> | <div><div></div></div> | HIGH      | ONN-Based         | <input checked="" type="checkbox"/> |
| CodeBERT      | <div><div></div></div> | <div><div></div></div> | HIGH      | Transformer-Based | <input type="checkbox"/>            |
| GraphCodeBERT | <div><div></div></div> | <div><div></div></div> | MEDIUM    | Transformer-Based | <input checked="" type="checkbox"/> |
| CodeTS        | <div><div></div></div> | <div><div></div></div> | MEDIUM    | Transformer-Based | <input type="checkbox"/>            |
| Devign        | <div><div></div></div> | <div><div></div></div> | LOW       | ONN-Based         | <input checked="" type="checkbox"/> |
| DfEPT         | <div><div></div></div> | <div><div></div></div> | VERY HIGH | ONN-Based         | <input checked="" type="checkbox"/> |

(3) 项目配置：包括黑名单配置和白名单配置，用户可以根据需要对 APK 进行分类和风险评级，同时还支持自定义名单配置，增加了管理的灵活性和精细度。

[illegible]

|           |   |
|-----------|---|
|           | <div><div>配置自定义名单</div><div><div><div><input type="checkbox"/> 涉诈</div><div><input type="checkbox"/> 涉赌</div><div><input type="checkbox"/> 涉黄</div><div><input type="checkbox"/> 其他黑灰产</div></div><div></div><div></div></div><div><div>添加黑名单</div><div>添加白名单</div></div></div>   |
| 问题        | <p>1.数据解析问题：</p> <p>（1）在使用 Androguard 对 APK 文件进行解析时，部分文件由于格式或内容问题导致解析失败，需要手动处理这些异常情况。</p> <p>（2）收集的 APK 文件中有一些标签不完整或不准确，需要进一步清洗和校正数据。</p> <p>2.大模型调用问题：</p> <p>（1）设计的大模型 Prompt 模板在调用时，生成的结果有时不够准确或详尽,需要进一步优化 Prompt 以提高模型的响应质量。</p> <p>（2）大模型的计算资源需求较高，在多次调用时容易造成系统负载，需要优化调用策略或增加计算资源。</p> <p>3.前端页面优化问题：</p> <p>（1）前端页面的响应速度在加载大量数据时较慢，尤其是在展示高危情况跟踪和风险 APP 总览时，需优化前端性能。</p> <p>（2）黑白名单的配置功能需要进一步完善，增加更多的分类和筛选条件，以满足不同场景的需求。</p>   |
| 下一步工作任务安排 | <p>1.数据清洗与解析优化：</p> <p>（1）继续清洗和校正已有的 APK 文件数据，确保标签的准确性和完整性。</p> <p>（2）优化 Androguard 的解析流程，处理异常文件的策略，确保所有 APK 文件都能成功解析并提取特征。</p> <p>2.分类模型训练：</p> <p>（1）基于已有的 APK 特征数据，训练多个分类模型（如使用 APK 特征分类和利用 APP 图标分类）。</p> <p>（2）使用不同的深度学习算法（如 Transformer、GNN 等）训练模型，以比较不同算法的性能。</p> <p>（3）评估各个模型的精度、速度和鲁棒性，选出表现最优的模型。</p> <p>3.大模型与多模态融合：</p> <p>（1）设计并实现一个 Agent 系统，能够调用多个小模型（如 APK 特征分类模型和 APP 图标分类模型）进行综合分析。</p> <p>（2）结合大模型和小模型的优点，实现多模态融合，提升整体检测系统的准确性和效率。</p> <p>4.前端页面优化：</p> <p>（1）优化前端数据加载和展示的性能，确保高危情况跟踪和风险 APP 总览的快速响应。</p> |

|                        |   |      |      |
|------------------------|---|------|------|
|                        | (2) 增加黑白名单配置的分类和筛选条件，提升用户操作的便捷性和灵活性。        |      |      |
| 项目进度自评<br>(在相应栏目内打“√”) | 正常  | 滞后延期 | 进展超前 |
|                        | √   |      |      |
| 教师意见                   | <div>教师签字：</div> <div>年      月      日</div> |      |      |

大数据与软件学院制表