Detection and Prevention of Android Ransomware

小标题



方案

1. 选择**CICDataset** 数据集

进行机器学习训练

模型就选择机器学习分类的模型，可能为了移动端部署，实现bp神经网络和常用饿分类的，选一个效果好的

数据集的介绍资料：<https://www.unb.ca/cic/datasets/invesandmal2019.html>

我会给论文老师先发一部分数据集样例

我们调试的是：

<https://keras.io/zh/getting-started/sequential-model-guide/>

<https://keras.io/zh/models/about-keras-models/>

理论部分：<https://zhuanlan.zhihu.com/p/487229356>

这个ANN模型（主要的）和其他传统机器学习分类模型或者算法（也许为了对比，加一些逻辑回归模型做比较，也是是其他算法，启发式的之类，这个对比的可以先待定，后续加上），先主要写ANN。

ANN：我们基于keras的Sequential模型可以构建非常复杂的神经网络，实现全连接神经网络：顺序模型是多个网络层的线性堆叠。通过将网络层实例的列表传递给 Sequential 的构造器，来创建一个 Sequential 模型。

模型需要知道它所期望的输入的尺寸，因此顺序模型中的第一层（且只有第一层，因为下面的层可以自动地推断尺寸）需要接收关于其输入尺寸的信息。有几种方法来做到这一点：

input\_shape：传递一个 input\_shape 参数给第一层。它是一个表示尺寸的元组 (一个整数或 None 的元组，其中 None 表示可能为任何正整数)。在 input\_shape 中不包含数据的 batch 大小。

input\_dim：某些 2D 层，例如 Dense，支持通过参数 input\_dim 指定输入尺寸，某些 3D 时序层支持 input\_dim 和 input\_length 参数。

batch\_size：如果你需要为你的输入指定一个固定的 batch 大小（这对 stateful RNNs 很有用），你可以传递一个 batch\_size 参数给一个层。如果你同时将 batch\_size=32 和 input\_shape=(6, 8) 传递给一个层，那么每一批输入的尺寸就为 (32，6，8)

对照模型：估计是LR/启发式算法/KNN/SVM之类的，这个可以以后加上

2. 做一个后端**api**服务，在一个后端**webserver**部署起来（**wsgiserver/flask api**等）

--- 整体的流程 ----

我们的前后端主要基于：flask+sqlchemy + numpy+ html5+vue+jquery技术栈

先把第一步训练好的机器学习模型和参数缓存起来，然后放在flask同级别地方方便调用。

对于科学数据集的管理：是flask 后端+ flask cros ,前端vue+bootstrap+ jinja 接受之后，传递到后端sqlchemy进行存储

整体上主要是从前端Jinja的template页面触发事件，然后通过Flask API传递到android 预测client调用

Flask 是一个基于 Python 的轻量级 Web 框架，WSGI 工具箱采用 Werkzeug，模板引擎使用 Jinja2。由于其不依赖于特殊的工具或库，并且没有数据抽象层、表单验证或是其他任何已有多种库可以胜任的功能，从而保持核心简单、易于扩展，而被定义为"微"框架。但是，Flask 可以通过扩展来添加应用功能。并且 Flask 具有自带开发用服务器和 debugger、集成单元测试和 RESTful 请求调度 (request dispatching)、支持 secure cookie 的特点。我们就主要使用Flask的网站部分和wsgi写API部分

# --- 分个介绍 ----

Jinja2 是基于 Python 的模版引擎，支持 Unicode，具有集成的沙箱执行环境并支持选择自动转义。Jinja2 拥有强大的自动 HTML 转移系统，可以有效的阻止跨站脚本攻击；通过模版继承机制，对所有模版使用相似布局；通过在第一次加载时将源码转化为 Python 字节码从而加快模版执行时间。我们的网站看得到的页面部分是这块开发。

Flask

1、Flask主要包括Werkzeug和Jinja2两个核心函数库，他们分别负责阢处理和安全方面的工鞥呢，这些基础函数为Web项目开发过程提供了丰富的基础组件。

　　2、Flask中的Jinja2模板引擎，提高了前端代码的复用率。可以大大提高开发效率并且有利于后期的开发与维护。

　　3、Flask不会指定数据库和模板引擎等对象，用户可以根据需要自己选择各种数据库。

　　4、Flask不提供表单验证功能，在项目实施过程中可以自由配置，从而为应用程序开发提供数据库抽象层基础组件，支持进行表单数据合法性验证、文件上传处理、用户身份认证和数据库集成等功能。

Flask的特点可以概括为：因为灵活，轻便高效，被业界所认可，同时拥有基于Werkzeug、Jinja2等一些开源库，拥有内置服务器和单元测试，适配RESTful。我们使用flask编写网站的用户登录/注册/权限管理/个人主页/机器学习训练和可视化的前后台逻辑部分，非常方便后续进行扩展。

我将使用 SQLite，这是一个小型 SQL 数据库实现，非常容易启动和运行。请记住，您可能想在生产环境中考虑更可靠的数据库，例如 PostgreSQL 或 MySQL。图像

flask\_sqlalchemy

要在 Flask 项目中设置 SQLAlchemy，我们可以导入 flask\_sqlalchemy 软件包（我们之前已安装），然后将 Flask app 变量包装在新的 SQLAlchemy 对象。我们还希望在 Flask 应用程序配置中设置 SQLALCHEMY\_DATABASE\_URI 以指定我们要使用的数据库以及如何访问它

# -- api 风格---

最后，我们可以开始定义 RESTful 处理程序。我们将使用 Flask-RESTful 软件包，这是一组工具，可帮助我们使用面向对象的设计来构建 RESTful 路由。

REST架构风格

六条设计规范定义了一个 REST 系统的特点:

客户端-服务器: 客户端和服务器之间隔离，服务器提供服务，客户端进行消费。

无状态: 从客户端到服务器的每个请求都必须包含理解请求所必需的信息。换句话说， 服务器不会存储客户端上一次请求的信息用来给下一次使用。

可缓存: 服务器必须明示客户端请求能否缓存。

分层系统: 客户端和服务器之间的通信应该以一种标准的方式，就是中间层代替服务器做出响应的时候，客户端不需要做任何变动。

统一的接口: 服务器和客户端的通信方法必须是统一的。

按需编码: 服务器可以提供可执行代码或脚本，为客户端在它们的环境中执行。这个约束是唯一一个是可选的。

Flask-RESTful

我们需要设置 Flask-RESTful 扩展名才能在 Flask 服务器中启动并运行。Flask-RESTful 是一个 Flask 扩展，它添加了快速构建 REST APIs 的支持。它当然也是一个能够跟你现有的ORM/库协同工作的轻量级的扩展。Flask-RESTful 鼓励以最小设置的最佳实践

Vue

整个平台的前端部分和可视化部分我们主要是使用vue+jquery+html5: Vue 是一套用于构建用户界面的渐进式 JavaScript 框架 ；同时它是一个典型的 MVVM 模型的框架（即：视图层-视图模型层-模型层）;HTML5是HTML的新标准，是一种超文本标记语言，是用来创建网页的标准标记语言，通过一系列的标识，来规范网络上的文档格式;区别：

        1.vue是一个渐进式 JavaScript 框架，而HTML5是一种超文本标记语言 2.在开发中vue框架通过mvvm的模式，解耦了视图层与模型层，而HTML5原生开中数据与标签紧耦合；    但是vue和html5可以进行结合:    vue是一个前端框架，但还是建立在HTML ，CSS ，JavaScript的基础之上的，通过编译之后依然是HTML+CSS+JavaScript组成。

3. 把第一步的机器学习模型训练好的模型缓存起来，拷贝到 **api**后端工程，集成在一起，提供接受输入，

模型进行分类预测的**api**

4. **android**手机端通过连线**ado** 方式 /（这块还在周末才能整个调完）

https://stackoverflow.com/questions/30173080/how-can-i-analyse-a-apk-file-in-android-application-get-the-permissions

写一个android app 即成这个overflow说的库，然后调用库分析某些本地库权限list/或者是分析api调用序列（看第一步使用的是什么数据集而定）

然后app端根据获取的权限，动态传回机器学习api接口，然后获得预测结果，进行不同的弹框提示

（这一块还待确认，因为overflow大家说某些机型这个获取别的apk的权限/api序列是不work的），