DEC 22ND, 2015

SUPOR: Precise and Scalable Sensitive User Input Detection for Android Apps

论文下载：http://www.cc.gatech.edu/~klu38/publications/security15.pdf

摘要

在本文中，作者提出了一种利用自然语言处理的方式处理android应用的UI界面，通过分析XML文件中相关部件的属性、描述字符、提示，并且结合相关加载UI的代码片段，判断出哪些控件是关于用户隐私信息输入的。

分析框架

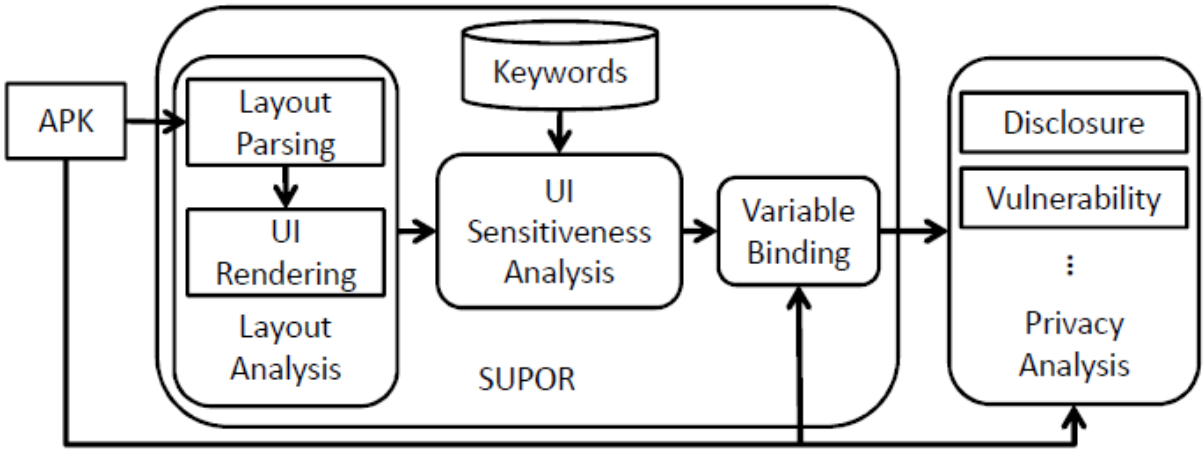


Figure 5: Overview of SUPOR.

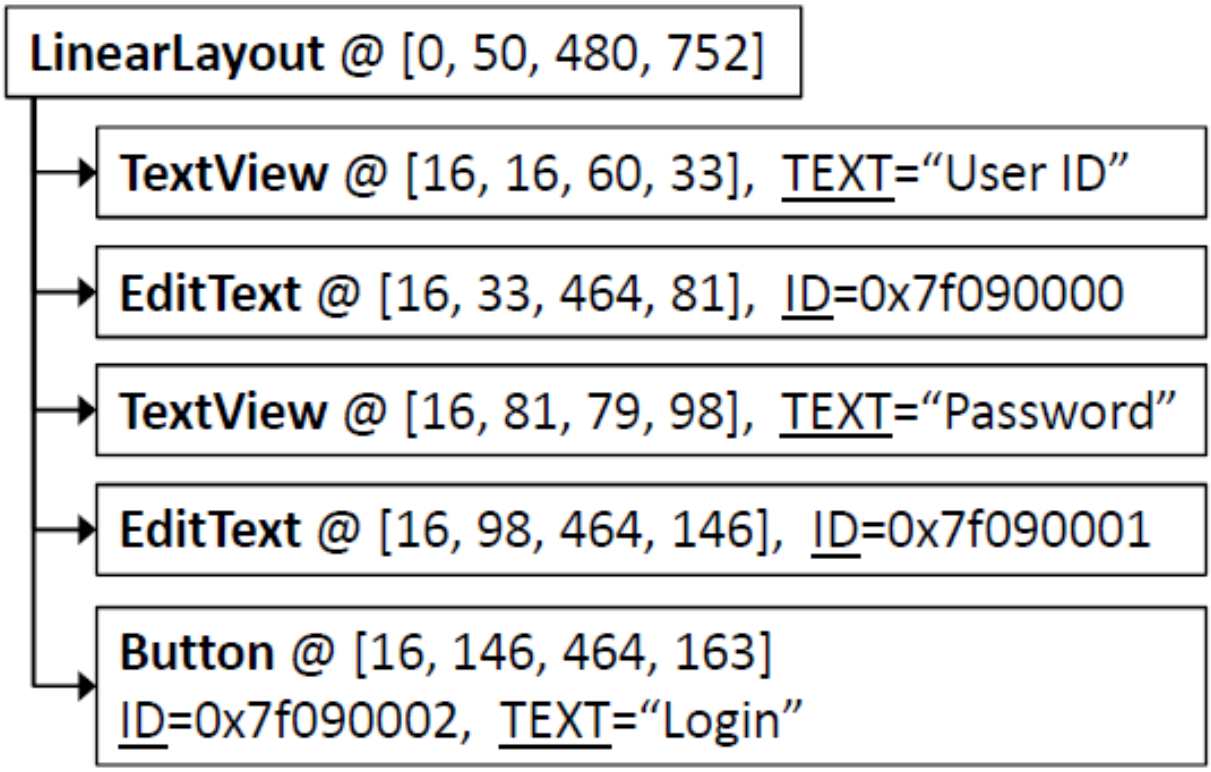
Layout分析

利用apktool提取出layout、string、和代码：

1 通过解析layout文件，鉴别出哪一个UI界面包含用户输入的控件。

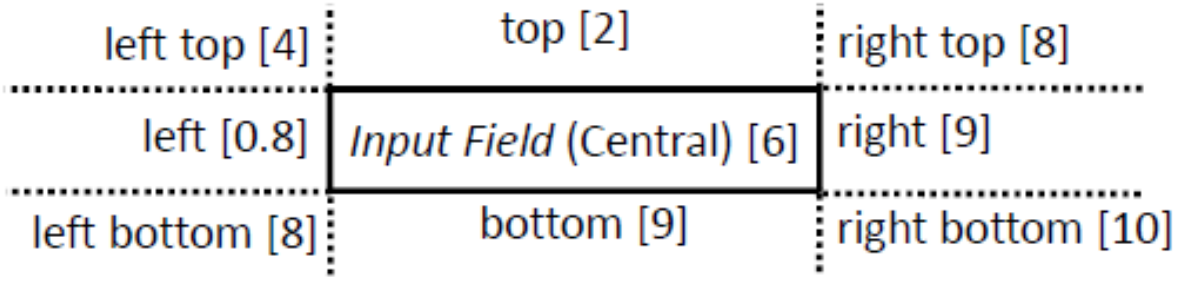
2 通过自然语言处理的方式将第一步获取的Layout处理成树状图。

3 根据xml计算出每一个控件的相对和绝对位置。



UI隐私分析

1 是否有textPassword等属性
2 判断有没有显而易见的hint
3 通过描述性的文件。



取位置权重*相对位置的最小值作为最匹配的描述性文件，在自己建立的关键字数据库中查找是否有匹配，如果有则表明对应的输入框是隐私输入。

变量绑定

- 通过上述方式找到相关控件和Layout的ID。在代码中查找绑定对应的ID的API(findViewById)
- 为了防止开发者在不同layout中设置相同的控件ID，在上述过程中找到的API做代码切片，往前搜索setContentView，减少误报。

关键字数据库

- 通过对下载的54371个ap的分析，将它们的资源文件经过NLP的处理，按照频率排序，在加入人工分析，组成关键字数据库。

实验

	#Apps	Percentage
Without Layout Files	625	3.91%
Without Input Fields	5,711	35.69%
Without Sensitive Input Fields	4,731	29.57%
With Sensitive Input Fields	4,922	30.76%
Parsing Errors	11	0.07%
TOTAL	16,000	100.00%

- 平均1分钟11.1个app（一共8台服务器集群）
- 1、挑选了20个没有隐私输入的app，发现1个误报：没有第三方库。2、20个有隐私输入的app，4FP和3FN。