(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108573151 A
(43)申请公布日 2018.09.25

(21)申请号 201710148636.4

(22)申请日 2017.03.10

(71)申请人 武汉安天信息技术有限责任公司
地址 430000 湖北省武汉市东湖新技术开发区软件园东路1号软件产业4-1期B4栋12层01室

(72)发明人 章康 冯泽 乔伟

(51)Int.Cl.
G06F 21/56(2013.01)

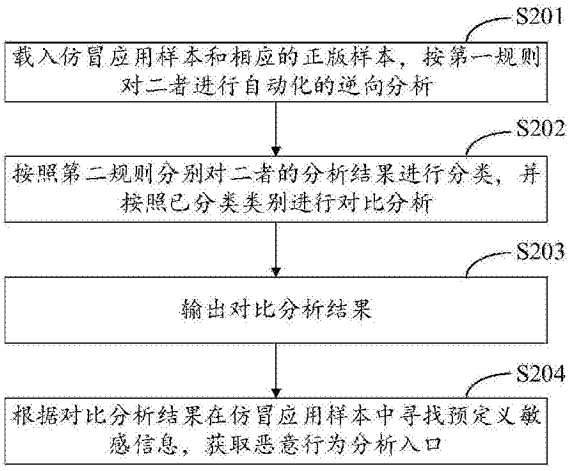
权利要求书2页 说明书4页 附图1页

(54)发明名称

一种仿冒应用分析系统及方法

(57)摘要

本发明提供了一种仿冒应用分析系统及方法，对仿冒应用样本及对应的正版应用样本进行自动化的逆向分析，然后对分析结果进行科学的分类对比，能够实现对仿冒应用更细粒度、更直观、更有效的分析，最后根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口，以实现仿冒应用恶意行为的分析，进而为用户提供完善的防护手段。



CN 108573151 A

1.一种仿冒应用分析系统，其特征在于，包括加载分析引擎、对比分析引擎、输出引擎和敏感信息分析引擎，其中：

加载分析引擎，用于载入仿冒应用样本和相应的正版样本，并按第一规则分别对仿冒应用样本和正版应用样本进行逆向分析，其中，所述第一规则为对样本从应用文件属性信息、应用的源码属性信息、应用的视图信息中的至少一个角度进行分析；

对比分析引擎，用于按照第二规则分别对仿冒应用样本和正版应用样本的分析结果进行分类，并对仿冒应用样本和相应的正版样本按照已分类类别进行对比分析；

输出引擎，用于输出对比分析结果；

敏感信息分析引擎，用于根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口。

2.如权利要求1所述的系统，其特征在于，所述第二规则包括对样本的分析结果按照文本类数据、树形结构数据、图形文件数据进行分类。

3.如权利要求2所述的系统，其特征在于，所述文本类数据包括：Java源码文本、Smali源码文本、AndroidManifest文本、资源文本、签名表格信息；树形结构数据包括：源码类树形结构数据、APK包树形结构数据；图形文件数据包括每个Smali与Java的CFG信息。

4.如权利要求1所述的系统，其特征在于，所述输出引擎还用于对得到对比分析结果区分显示，其中，所述区分方法包括：利用颜色、注释、字体进行区分。

5.一种仿冒应用分析方法，其特征在于，包括以下步骤：

载入仿冒应用样本和相应的正版样本，按第一规则分别对仿冒应用样本和正版应用样本进行逆向分析，其中，所述第一规则为对样本从应用文件属性信息、应用的源码属性信息、应用的视图信息中的至少一个角度进行分析；

按照第二规则分别对仿冒应用样本和正版应用样本的分析结果进行分类，并对仿冒应用样本和相应的正版样本按照已分类类别进行对比分析；

输出对比分析结果；

根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口。

6.如权利要求5所述的方法，其特征在于，所述第二规则包括对样本的分析结果按照文本类数据、树形结构数据、图形文件数据进行分类。

7.如权利要求5所述的方法，其特征在于，所述文本类数据包括：Java源码文本、Smali源码文本、AndroidManifest文本、资源文本、签名表格信息；树形结构数据包括：源码类树形结构数据、APK包树形结构数据；图形文件数据包括每个Smali与Java的CFG信息。

8.如权利要求5所述的方法，其特征在于，输出对比分析结果时还对得到对比分析结果区分显示，其中，所述区分方法包括：利用颜色、注释、字体进行区分。

9.如权利要求1所述的系统或如权利要求5所述的方法，其特征在于，所述预定义敏感信息包括：Smali源码中的手机号、Java源码中的手机号、链接网络信息、AndroidManifest.xml中的特殊权限及行为。

10.如权利要求1所述的系统或如权利要求5所述的方法，其特征在于，所述应用文件属性信息包括：应用文件大小、应用文件Hash值；所述应用的源码属性信息包括：应用文件字符串信息、Java类源码结构、Smali转java的反编译信息；所述应用的视图信息包括：每个Smali与Java的CFG信息、应用文件的签名信息、资源文件信息、应用程序图标、应用文件结

构信息、AndroidManifest解码信息。

一种仿冒应用分析系统及方法

技术领域

[0001] 本发明涉及移动安全技术领域，尤其涉及一种仿冒应用分析系统及方法。

背景技术

[0002] 随着移动APP的不断丰富和发展，仿冒应用产业链的规模也迅速扩大。据官方公布的2015年度仿冒应用统计报告，95%的热门APP被仿冒应用困扰。因此，鉴别和分析正版应用和仿冒应用是解决移动APP安全的一大重要问题。

[0003] 目前针对仿冒应用的分析多数采用二进制对比的方法(例如，CFG控制流程图的对比)，虽然二进制对比正版应用和仿冒应用能够判断出应用是否是仿冒应用，能够标记哪里被修改，但是无法对比分析出正版应用和仿冒应用更加细化的信息(例如，应用的源码被修改的内容)，更加无法分析出仿冒应用的恶意行为，进而无法给用户提供完善的防护手段。

发明内容

[0004] 针对上述技术问题，本发明提供了一种仿冒应用分析系统及方法，能够得到正版应用和仿冒应用更细粒度的分析结果，并能分析出仿冒应用的恶意行为，进而为用户提供完善的防护手段。

[0005] 本发明公开的仿冒应用分析系统，包括加载分析引擎、对比分析引擎、输出引擎和敏感信息分析引擎，其中：

加载分析引擎，用于载入仿冒应用样本和相应的正版样本，并按第一规则分别对仿冒应用样本和正版应用样本进行逆向分析，其中，所述第一规则为对样本从应用文件属性信息、应用的源码属性信息、应用的视图信息中的至少一个角度进行分析；

对比分析引擎，用于按照第二规则分别对仿冒应用样本和正版应用样本的分析结果进行分类，并对仿冒应用样本和相应的正版样本按照已分类类别进行对比分析；

输出引擎，用于输出对比分析结果；

敏感信息分析引擎，用于根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口。

[0006] 进一步的，所述第二规则包括对样本的分析结果按照文本类数据、树形结构数据、图形文件数据进行分类。

[0007] 进一步的，所述文本类数据包括：Java源码文本、Smali源码文本、AndroidManifest文本、资源文本、签名表格信息；树形结构数据包括：源码类树形结构数据、APK包树形结构数据；图形文件数据包括每个Smali与Java的CFG信息。

[0008] 进一步的，所述输出引擎还用于对得到对比分析结果区分显示，其中，所述区分方法包括：利用颜色、注释、字体大小进行区分。

[0009] 进一步的，所述预定义敏感信息包括：Smali源码中的手机号、Java源码中的手机号、链接网络信息、AndroidManifest.xml中的特殊权限及行为。

[0010] 进一步的，所述应用文件属性信息包括：应用文件大小、应用文件Hash值；所述应

用的源码属性信息包括:应用文件字符串信息、Java类源码结构、Smali转java的反编译信息;所述应用的视图信息包括:每个Smali与Java的CFG信息、应用文件的签名信息、资源文件信息、应用程序图标、应用文件结构信息、AndroidManifest解码信息。

[0011] 本发明还公开了一种仿冒应用分析方法,包括以下步骤:

载入仿冒应用样本和相应的正版样本,并按第一规则分别对仿冒应用样本和正版应用样本进行逆向分析,其中,所述第一规则为对样本从应用文件属性信息、应用的源码属性信息、应用的视图信息中的至少一个角度进行分析;

按照第二规则分别对仿冒应用样本和正版应用样本的分析结果进行分类,并对仿冒应用样本和相应的正版样本按照已分类类别进行对比分析;

输出对比分析结果;

根据对比分析结果在仿冒应用样本中寻找预定义敏感信息,获取恶意行为分析入口。

[0012] 进一步的,所述第二规则包括对样本的分析结果按照文本类数据、树形结构数据、图形文件数据进行分类。

[0013] 进一步的,所述文本类数据包括:Java源码文本、Smali源码文本、AndroidManifest文本、资源文本、签名表格信息;树形结构数据包括:源码类树形结构数据、APK包树形结构数据; 图形文件数据包括每个Smali与Java的CFG信息。

[0014] 进一步的,输出对比分析结果时还对得到对比分析结果区分显示,其中,所述区分方法包括:利用颜色、注释、字体大小进行区分。

[0015] 进一步的,所述预定义敏感信息包括:Smali源码中的手机号、Java源码中的手机号、链接网络信息、AndroidManifest.xml中的特殊权限及行为。

[0016] 进一步的,所述应用文件属性信息包括:应用文件大小、应用文件Hash值;所述应用的源码属性信息包括:应用文件字符串信息、Java类源码结构、Smali转java的反编译信息;所述应用的视图信息包括:每个Smali与Java的CFG信息、应用文件的签名信息、资源文件信息、应用程序图标、应用文件结构信息、AndroidManifest解码信息。

[0017] 本发明的有益效果是:

本发明对仿冒应用样本及对应的正版应用样本进行自动化的逆向分析,然后对分析结果进行科学的分类对比,能够实现对仿冒应用更细粒度、更直观、更有效的分析,最后根据对比分析结果在仿冒应用样本中寻找预定义敏感信息,获取恶意行为分析入口,以实现仿冒应用恶意行为的分析,进而为用户提供完善的防护手段。

[0018]

附图说明

[0019] 为了更清楚地说明本发明或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0020] 图1为本发明一种仿冒应用分析系统的结构示意图;

图2为本发明一种仿冒应用分析方法的流程图。

[0021]

具体实施方式

[0022] 为了使本技术领域的人员更好地理解本发明实施例中的技术方案，并使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图对本发明中技术方案作进一步详细的说明。

[0023] 本发明给出了一种仿冒应用分析系统的实施例，如图1所示，该分析系统包括加载分析引擎101、对比分析引擎102、输出引擎103和敏感信息分析引擎104，其中：

加载分析引擎101，用于载入仿冒应用样本和相应的正版样本，并按第一规则对二者进行自动化的逆向分析。

[0024] 具体的，所述第一规则为对样本从应用文件属性信息、应用的源码属性信息、应用的视图信息中的至少一个角度进行分析。

[0025] 可以理解的，应用文件属性信息可以应用文件大小、对文件做唯一标识的应用文件Hash值(如MD5值)等等。

[0026] 应用的源码属性信息可以包括应用文件字符串信息、Java类源码结构、Smali转java的反编译信息等等。

[0027] 应用的视图信息可以包括每个Smali与Java的CFG信息、应用文件的签名信息、资源文件信息、应用程序图标、应用文件结构信息、AndroidManifest解码信息等等。

[0028] 对比分析引擎102，用于按照第二规则分别对仿冒应用样本和正版应用样本的分析结果进行分类，并对仿冒应用样本和相应的正版样本按照已分类类别进行对比分析。

[0029] 输出引擎103，用于输出对比分析结果。

[0030] 为了对样本的分析结果进行更科学、直观的分类对比，第二规则可以为将样本的分析结果按照文本类数据、树形结构数据、图形文件数据三类进行划分。

[0031] (1)文本类数据包括：Java源码文本、Smali源码文本、AndroidManifest文本、资源文本、签名表格信息等等。

[0032] (2)树形结构数据包括：源码类树形结构数据、APK包树形结构数据等等。

[0033] (3)图形文件数据包括：每个Smali与Java的CFG信息等等

当然，也可以根据输出引擎103的实际情况对样本分析结果进行分类对比。

[0034] 为了能更加清楚、直观的了解对比分析结果，该输出引擎103还用于对得到对比分析结果区分显示，比如利用颜色、注释、字体大小等对分析结果进行区分。

[0035] 敏感信息分析引擎104，用于根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口。预定义敏感信息可以包括Smali源码中的手机号、Java源码中的手机号、链接网络信息、AndroidManifest.xml中的特殊权限及行为等等。利用敏感信息分析引擎104能获取恶意行为分析入口，以实现仿冒应用恶意行为的分析，进而为用户提供完善的防护手段。

[0036] 传统的仿冒应用的检测都是基于图标相似度、代码结构相似度等，准确率不高，而且也不能进行深入分析。本专利不仅仅可以通过布局相似性进行仿冒的判定，而且还提供人性化的界面辅助深入分析。

[0037] 另外，本发明还给出了一种仿冒应用分析方法的实施例，如图2所示，该分析方法

包括：

S201：载入仿冒应用样本和相应的正版样本，按第一规则对二者进行自动化的逆向分析。

[0038] 具体的，通常情况下样本文件为APK文件；第一规则为对样本从应用文件属性信息、应用的源码属性信息、应用的视图信息中的至少一个角度进行分析。

[0039] 可以理解的，应用文件属性信息可以应用文件大小、对文件做唯一标识的应用文件Hash值(如MD5值)等等。

[0040] 应用的源码属性信息可以包括应用文件字符串信息、Java类源码结构、Smali转java的反编译信息等等。

[0041] 应用的视图信息可以包括每个Smali与Java的CFG信息、应用文件的签名信息、资源文件信息、应用程序图标、应用文件结构信息、AndroidManifest解码信息等等。

[0042] S202：按照第二规则分别对仿冒应用样本和正版应用样本的分析结果进行分类，并对仿冒应用样本和相应的正版样本按照已分类类别进行对比分析。

[0043] S203：输出对比分析结果。

[0044] 为了对样本的分析结果进行更科学、直观的分类对比，第二规则可以为将样本的分析结果按照文本类数据、树形结构数据、图形文件数据三类进行划分。

[0045] (1)文本类数据包括：Java源码文本、Smali源码文本、AndroidManifest文本、资源文本、签名表格信息等等。

[0046] (2)树形结构数据包括：源码类树形结构数据、APK包树形结构数据等等。

[0047] (3)图形文件数据包括：每个Smali与Java的CFG信息等等

当然，也可以根据S203的实际情况对样本分析结果进行分类对比。

[0048] 为了能更加清楚、直观的了解对比分析结果，S203中还对得到对比分析结果区分显示，比如利用颜色、注释、字体大小等对分析结果进行区分。

[0049] 对比分析结果中的区别信息进行区分；其中，区分方法包括：利用颜色进行区分、利用注释进行区分、利用字体进行区分。

[0050] S204：根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口。

[0051] 预定义敏感信息可以包括Smali源码中的手机号、Java源码中的手机号、链接网络信息、AndroidManifest.xml中的特殊权限及行为等等。

[0052] 本发明对仿冒应用样本及对应的正版应用样本进行自动化的逆向分析，然后对分析结果进行科学的分类对比，能够实现对仿冒应用更细粒度、更直观、更有效的分析，最后根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口，以实现仿冒应用恶意行为的分析，进而为用户提供完善的防护手段。

[0053] 虽然通过实施例描绘了本发明，本领域普通技术人员知道，本发明有许多变形和变化而不脱离本发明的精神，希望所附的权利要求包括这些变形和变化而不脱离本发明的精神。

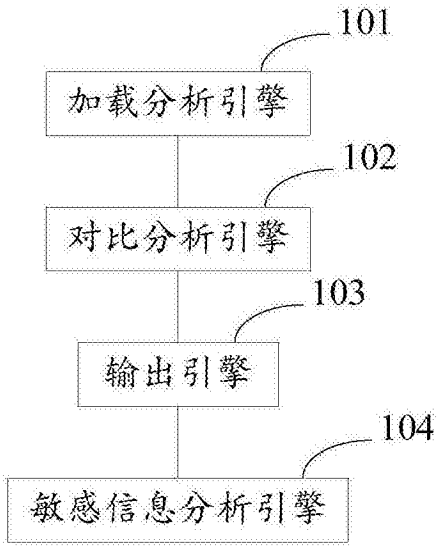

101
加载分析引擎
102
对比分析引擎
103
输出引擎
104
敏感信息分析引擎

图1

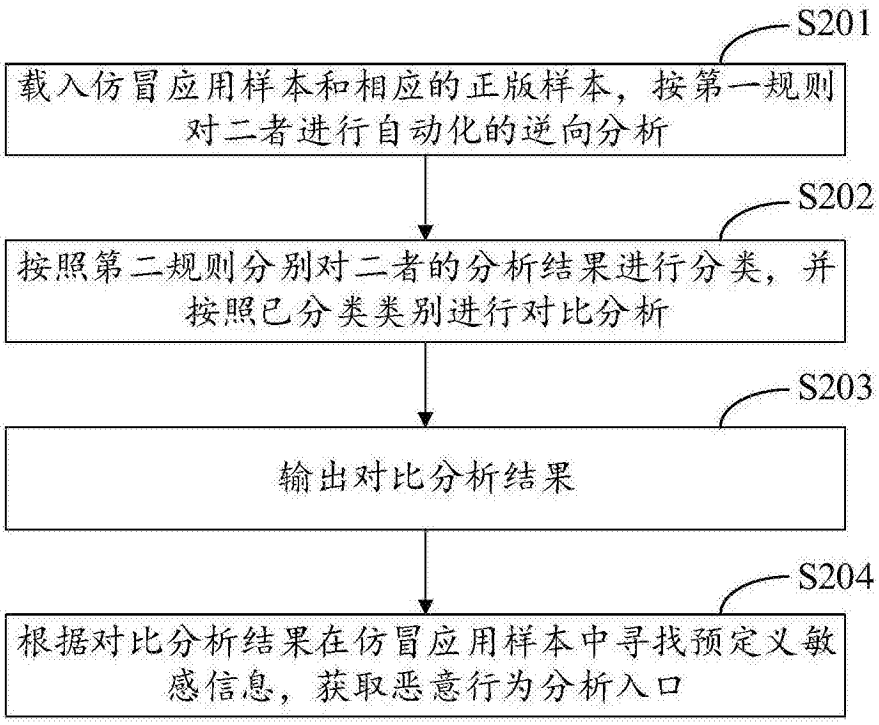

S201
载入仿冒应用样本和相应的正版样本，按第一规则对二者进行自动化的逆向分析
S202
按照第二规则分别对二者的分析结果进行分类，并按照已分类类别进行对比分析
S203
输出对比分析结果
S204
根据对比分析结果在仿冒应用样本中寻找预定义敏感信息，获取恶意行为分析入口

图2