

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 3 月 15 日 (15.03.2018)



(10) 国际公布号
WO 2018/045489 A1

- (51) 国际专利分类号:
G06F 9/44 (2006.01) **G06F 17/40** (2006.01)
- (21) 国际申请号: PCT/CN2016/098240
- (22) 国际申请日: 2016 年 9 月 6 日 (06.09.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 张庆龙 (ZHANG, Qinglong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (74) 代理人: 北京同达信恒知识产权代理有限公司 (TDIP & PARTNERS); 中国北京市海淀区知春路 7 号致真大厦 A1304-05 室, Beijing 100191 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

(54) Title: DATA COLLECTION METHOD, APPARATUS AND SYSTEM

(54) 发明名称: 一种数据采集方法、装置及系统

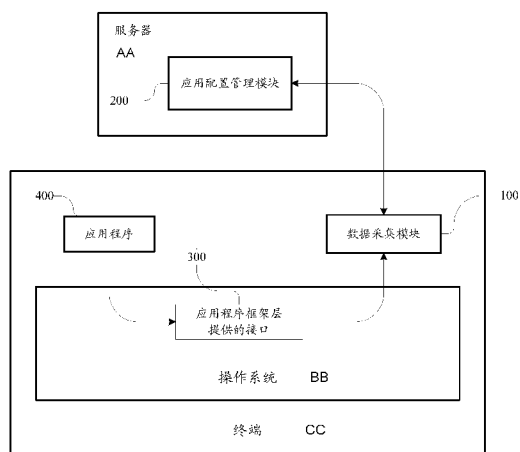


图 3

- 100 DATA COLLECTION MODULE
200 APPLICATION PROGRAM CONFIGURATION MANAGEMENT MODULE
300 INTERFACE PROVIDED BY A FRAMEWORK LAYER OF AN APPLICATION PROGRAM
400 APPLICATION PROGRAM
AA SERVER
BB OPERATING SYSTEM
CC TERMINAL

(57) Abstract: A data collection method, apparatus and system, for use in reducing the coupling degree between an application program and operations about processing and configuration related to data collection. In the present invention, information about an application program is obtained according to a framework layer of the application program of an operating system; data collection configuration information of the application program is queried according to the information, so as to obtain the data collection configuration information of the application program; and application data is collected according to the data collection configuration information of the application program and an interface provided by the framework layer of the application program of the operating system, the information being used for identifying the application program.

(57) 摘要: 一种数据采集方法、装置及系统, 用以降低数据采集相关处理和配置操作与应用程序的耦合度。本发明中, 基于操作系统应用程序框架层获取应用程序的信息, 根据所述信息查询应用程序数据采集配置信息, 得到应用程序的数据采集配置信息, 根据所述应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述应用程序的应用数据, 其中, 所述信息用于识别所述应用程序。

本国际公布：

- 包括国际检索报告(条约第21条(3))。

一种数据采集方法、装置及系统

技术领域

本发明涉及通信技术领域，尤其涉及一种数据采集方法、装置及系统。

背景技术

随着智能终端的不断普及，智能终端已经广泛嵌入到各种不同的生活场景中，包括聚会、购物、出行、娱乐、社交等等一系列社会活动。由此用户需在智能终端设备中安装使用各种应用程序，以满足上述社会活动。用户在使用终端设备时，产生大量与自己个体相关的数据，并且这些数据是用户属性的最直接体现。在这种情况下，单个应用中的数据难以刻画用户属性，如何整合智能终端设备上所有应用的有价值数据是一个待解决的问题。

目前，对终端进行数据采集可以分为两类，一类是应用内部集成采集工具包，采集工具包将数据上报服务器，若要采集多个应用的数据，则每个应用均要集成采集工具包，如图 1 所示。一类是直接在服务器端采集终端设备与服务器的交互日志。第二类采集方式相对第一类粒度较粗，比如采集用户性别，第一类采集方法是在用户编辑该属性的时候触发采集工具包中的配置项才能采集到，而第二类采集方法很难采集到用户敏感数据。

应用开发者针对不同应用程序的不同采集内容，定制不同的采集工具包，在开发的应用程序中集成为该应用程序定制的采集工具包，针对需要采集的页面配置应用数据采集项。在数据采集时，终端将采集到的数据上报服务器，在服务器端完成数据分析与数据存储。

可以看出，使用采集工具包对应用程序进行数据采集的方案中，采集工具包与应用程序之间紧耦合，这样可能带来以下问题：应用程序版本更新时，采集工具包也需要同步更新，使得应用程序开发维护复杂度较高。

因此，如何降低数据采集的相关处理和配置操作与应用程序之间的耦合度，是业界所亟待研究和解决的问题。

发明内容

本发明实施例提供一种数据采集方法、装置及系统，用以降低数据采集的相关处理和配置操作与应用程序的耦合度。

本发明实施例提供的数据采集方法，包括：

基于操作系统应用程序框架层获取第一应用程序的第一信息，根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，其中，所述第一信息用于识别所述第一应用程序；

基于操作系统应用程序框架层获取第二应用程序的第二信息，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据，其中，所述第二信息用于识别所述第二应用程序。

可选地，基于操作系统应用程序框架层获取第一应用程序的第一信息，包括：操作系统应用程序框架层中的应用管理器在第一应用程序激活时发送所述第一应用程序的第一信息，接收所述应用管理器发送的所述第一信息；基于操作系统应用程序框架层获取第二应用程序的第二信息，包括：操作系统应用程序框架层中的应用管理器在第二应用程序激活时发送所述第二应用程序的第二信息，接收所述应用管理器发送的所述第二信息。

可选地，所述应用管理器在第一应用程序调用重加载方法时，确定所述第一应用程序激活；所述应用管理器在第二应用程序的进程调用重加载方法时，确定所述第二应用程序激活。

可选地，所述第一信息包括以下信息中的一种或多种：第一应用程序的包名、第一应用程序的统一资源标识符 URI、第一应用程序对应的进程名、第一应用程序的版本；所述第二信息包括以下信息中的一种或多种：第二应用程序的包名、第二应用程序的 URI、第二应用程序对应的进程名、第二应

用程序的版本。

可选地，所述第一信息还包括：第一应用程序的类名或线程名；所述第二信息还包括：第二应用程序的类名或线程名。

可选地，所述应用程序数据采集配置信息包括：配置为需要进行数据采集的应用程序列表；需要进行数据采集的应用程序的数据采集项配置信息，一个应用程序的数据采集项配置信息用于定义对该应用程序的哪些应用数据进行采集。

可选地，所述数据采集项配置信息包括以下信息中的一种或多种：应用程序控件的名称、类型、对应的对象名称中的一种或多种；用于网页的标记性语言中的标签名称、标签类型中的一种或多种。

可选地，采集所述第一应用程序的应用数据之后，还包括：将采集到的第一应用程序的应用数据保存到本地数据库；采集所述第二应用程序的应用数据之后，还包括：将采集到的第二应用程序的应用数据保存到本地数据库。

可选地，将采集到的第一应用程序的应用数据保存到本地数据库之前，还包括：根据采集到的第一应用程序的数据所归属的属性类别，对采集到的第一应用程序的数据进行分类；将采集到的第二应用程序的应用数据保存到本地数据库之前，还包括：根据采集到的第二应用程序的数据所归属的属性类别，对采集到的第二应用程序的数据进行分类。

可选地，还包括：基于操作系统框架层确定有应用程序被安装或有应用程序被更新时，向服务器发送获取请求，所述获取请求用于请求获取所述被安装或所述被更新的应用程序的数据采集配置信息；接收所述服务器根据所述获取请求返回的数据采集配置信息。

本发明实施例提供的数据采集装置，包括：

应用信息获取单元，用于基于操作系统应用程序框架层获取第一应用程序的第一信息，以及基于操作系统应用程序框架层获取第二应用程序的第二信息；

查询单元，用于根据所述第一信息查询应用程序数据采集配置信息，得

到第一应用程序的数据采集配置信息，以及，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息；其中，所述第一信息用于识别所述第一应用程序，所述第二信息用于识别所述第二应用程序；

数据采集单元，用于根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，以及，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据。

可选地，所述应用信息获取单元具体用于：接收应用程序框架层中的应用管理器发送的所述第一信息，其中所述第一信息是所述应用管理器在第一应用程序激活时发送的；以及，接收应用程序框架层中的应用管理器发送的所述第二信息，其中所述第二信息是所述应用管理器在第二应用程序激活时发送的。

可选地，所述应用管理器在第一应用程序调用重加载方法时，确定所述第一应用程序激活；所述应用管理器在第二应用程序的进程调用重加载方法时，确定所述第二应用程序激活。

可选地，所述第一信息包括以下信息中的一种或多种：第一应用程序的包名、第一应用程序的统一资源标识符 URI、第一应用程序对应的进程名、第一应用程序的版本；所述第二信息包括以下信息中的一种或多种：第二应用程序的包名、第二应用程序的 URI、第二应用程序对应的进程名、第二应用程序的版本。

可选地，所述第一信息还包括：第一应用程序的类名或线程名；所述第二信息还包括：第二应用程序的类名或线程名。

可选地，所述应用程序数据采集配置信息包括：配置为需要进行数据采集的应用程序列表；需要进行数据采集的应用程序的数据采集项配置信息，一个应用程序的数据采集项配置信息用于定义对该应用程序的哪些应用数据进行采集。

可选地，所述数据采集项配置信息包括以下信息中的一种或多种：应用程序控件的名称、类型、对应的对象名称中的一种或多种；用于网页的标记性语言中的标签名称、标签类型中的一种或多种。

可选地，还包括：数据存储单元，用于在所述数据采集单元采集所述第一应用程序的应用数据之后，将采集到的第一应用程序的应用数据保存到本地数据库；以及，在所述数据采集单元采集所述第二应用程序的应用数据之后，将采集到的第二应用程序的应用数据保存到本地数据库。

可选地，还包括：数据分类单元，用于在所述数据存储单元将采集到的第一应用程序的应用数据保存到本地数据库之前，根据采集到的第一应用程序的数据所归属的属性类别，对采集到的第一应用程序的数据进行分类；以及，在所述数据存储单元将采集到的第二应用程序的应用数据保存到本地数据库之前，根据采集到的第二应用程序的数据所归属的属性类别，对采集到的第二应用程序的数据进行分类。

可选地，还包括：配置单元，用于基于操作系统框架层确定有应用程序被安装或有应用程序被更新时，向服务器发送获取请求，所述获取请求用于请求获取所述被安装或所述被更新的应用程序的数据采集配置信息；以及，接收所述服务器根据所述获取请求返回的数据采集配置信息。

本发明实施例提供的数据采集系统，包括：

服务器，用于为终端配置数据采集配置信息；

终端，用于基于操作系统应用程序框架层获取第一应用程序的第一信息，根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，其中，所述第一信息用于识别所述第一应用程序；以及，基于操作系统应用程序框架层获取第二应用程序的第二信息，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第

二应用程序的应用数据，其中，所述第二信息用于识别所述第二应用程序。

本发明实施例提供的通信装置，包括：处理单元和存储器；

所述存储器，用于存储计算机程序指令；

所述处理单元，耦合到所述存储器，用于读取所述存储器存储的计算机程序指令，并执行上述数据采集方法。

本发明的上述实施例中，对于第一应用程序，可基于操作系统应用程序框架层获取第一应用程序的第一信息，根据第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集第一应用程序的应用数据，对于其他应用程序，比如第二应用程序，也可采用上述方式进行数据采集。在操作系统中，应用程序框架层用于对运行于上的各种应用程序进行管理，因此基于应用程序框架层可获得运行于上的应用程序的相关信息；又由于应用程序的应用数据是基于应用程序框架层提供的接口采集的，该接口为系统级接口，不依赖于应用程序，可基于该接口对不同的应用程序进行数据采集，与现有技术相比，无需针对不同的应用程序制定不同的采集工具包，也无需将采集工具包集成于应用程序，从而降低了数据采集的相关处理和配置操作与应用程序的耦合度。

附图说明

图 1 为现有技术中的应用程序与采集工具包之间紧耦合的示意图；

图 2 为本发明实施例中基于 Android 平台的移动操作系统的架构示意图；

图 3 为本发明实施例提供的用于数据采集的系统架构示意图；

图 4 为本发明实施例提供的数据采集流程示意图；

图 5 为本发明实施例中的数据采集模块的结构示意图；

图 6 为本发明实施例中的数据采集单元的数据采集过程示意图；

图 7 为本发明实施例中的应用配置管理模块的处理流程示意图；

图 8 为本发明实施例中以 Android 系统为例的基于应用程序框架层提供的

接口实现的数据采集时序图；

图 9 为本发明实施例中应用程序数据采集配置信息内容的示意图；

图 10 为本发明实施例中应用程序用户界面的示意图；

图 11 为本发明实施例中应用程序更新前后的用户界面示意图；

图 12 为本发明实施例中应用程序数据采集配置信息更新前后的示意图；

图 13 为本发明实施例提供的数据采集装置的结构示意图；

图 14 为本发明实施例提供的数据采集装置的结构示意图。

具体实施方式

为了使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明作进一步地详细描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例，都属于本发明保护的范围。

本发明实施例中，在终端的操作系统的应用程序框架层中提供有应用程序数据采集接口，该接口用于提供对运行于操作系统应用程序框架层的应用程序的数据采集功能。在采集应用程序的应用数据时，基于该接口采集该应用程序的应用数据。由于该接口由操作系统应用程序框架层提供，因此基于接口可采集运行于操作系统应用程序框架的不同应用程序，获得这些应用程序的应用数据，与现有技术相比，无需针对不同的应用程序制定不同的采集工具包，也无需将采集工具包集成于应用程序，从而降低了数据采集的相关处理和配置操作与应用程序的耦合度。

其中，所述终端可包括多种类型的终端，比如可以包括移动终端，如手机、PDA（Personal Digital Assistant，掌上电脑）、智能终端、车载终端或智能穿戴设备等。这里的终端泛指能够运行应用程序的电子通信设备。

其中，本发明实施例对终端所使用的操作系统的种类不作限制。比如，该操作系统可以是基于 Android 平台的移动操作系统，也可以是基于云平台的

移动操作系统（如云操作系统），或者其他操作系统。

图 2 示出了基于 Android 平台的移动操作系统的架构示意图。该移动操作系统的软件层次结构自下而上分为以下几个层次：第一、操作系统层（OS），该层主要包括操作系统内核，比如该层可包括各种驱动（比如显示驱动、闪存驱动等）以及电源管理等；第二、各种库（Libraries）和 Android 运行环境（Android RunTime）；第三、应用程序框架（Application Framework），该层中可包括各种管理服务，比如诸如应用管理器（Activity Manager）、窗口管理器（windows Manager）、资源管理器（Resource Manager）、位置管理器（Location Manager）等服务；第四、应用程序（Application）。

应用程序框架层为应用程序的开发者提供多种应用程序编程接口（Application Programming Interface，简称 API），它实际上是一个应用程序的框架。由于上一层的应用程序是以 JAVA 构建的，因此本层提供的 APIs，包含了用户界面（User Interface，简称 UI）程序中所需要的各种控件，比如与本发明实施例相关的控件包括：列表（Lists）、栅格（Grids）、文本框（Text boxes）、按钮（Buttons）等视图组件（Views），或者嵌入式的 Web 浏览器等。本发明实施例中的应用程序数据采集接口可以是基于应用程序框架层的接口，即由应用程序框架层提供应用程序数据采集接口。

以上虽然是以基于 Android 平台的操作系统为例描述，但对于其他类型的操作系统，也可类似地由操作系统中的应用程序框架层提供用于对应用程序进行访问的应用程序数据采集接口。

图 3 示出了本发明实施例提供的用于数据采集的系统架构示意图。如图所示，该系统架构中可包括：数据采集模块 100、应用配置管理模块 200，应用程序框架层提供的接口 300、应用程序 400。其中，应用配置管理模块 200 为可选模块。

其中，数据采集模块 100 和应用程序框架层提供的接口 300 位于终端，应用配置管理模块 200 位于服务器。终端和服务器之间可通过网络进行通信，终端中的数据采集模块 100，与服务器中的应用配置管理模块 200 可进行通信。

所述网络可包括移动网络，比如蜂窝移动通信网络或者是采用其他移动通信技术的网络。

本发明实施例中，应用程序可运行于操作系统提供的应用程序框架，有自己的运行空间，进程间互不干扰，但共同点是均可被基于应用程序框架层提供的接口 300 提供的接口进行访问。

应用程序框架层可提供多种接口，这其中可包括应用框架层中现有的系统服务（比如应用管理器）所提供的接口，还包括本发明实施例所提供的数据采集接口。数据采集接口主要用于提供对运行于操作系统应用程序框架层的应用程序进行数据采集的功能，数据采集模块 100 可基于该接口，获得应用程序 400 在运行过程中产生的应用数据（比如对于在线交易类的应用程序来说，该类应用程序的应用数据可包括交易对象的类型、交易额等）。

应用配置管理模块 200 可将数据采集配置信息给数据采集模块 100，以使数据采集模块 100 根据该数据采集配置信息采集应用程序的应用数据。其中，数据采集配置信息用于定义哪些应用程序需要进行数据采集，和/或，针对需要进行数据采集的应用程序，需要采集哪种应用数据等。

应用程序框架层提供的接口 300 可作为操作系统的一部分，在操作系统安装时被安装到终端中。数据采集模块 100 可作为独立的应用程序（比如可以是系统应用程序）或服务（比如系统服务）进行安装，在安装过程中可对数据采集配置信息进行配置。数据采集配置信息可作为数据采集模块 100 的安装包的组成部分，在数据采集模块 100 安装的过程中进行配置，也可在安装过程中或安装完成后，通过与服务器中的应用配置管理模块 200 进行交互，以获得数据采集配置信息进行配置。

基于上述系统架构，图 4 示出了本发明实施例提供的数据采集流程，该流程可由数据采集模块执行，该流程可包括如下步骤：

步骤 401: 基于操作系统应用程序框架层，获取第一应用程序的第一信息。

该步骤中，操作系统应用程序框架层中的应用管理器可在第一应用程序激活时，向数据采集模块发送第一应用程序的第一信息。其中，应用管理器

可对应用程序的生命周期、应用程序的状态等进行管理，它可提供多种用于实现应用程序管理功能的接口，调用这些接口提供的方法可实现应用程序管理功能。比如，当第一应用程序调用应用管理器提供的重加载方法（如 onResume 方法）时，应用管理器可确定第一应用程序激活，进而向数据采集模块发送第一应用程序的第一信息。

其中，第一信息用于识别第一应用程序。作为应用程序的识别信息，第一信息中至少包括第一应用程序的标识信息，该标识信息用于唯一标识一个应用程序，比如该标识信息可以是应用程序的包名或统一资源定位符（Uniform Resource Identifier，简称 URI）或对应的进程名（通常情况下，一个应用程序对应于一个进程）。一个应用程序可包含多个线程，因此进一步地，第一信息中还可包括第一应用程序中的线程的名称。以视频播放应用程序为例，该应用程序中可包括音乐播放线程和音乐下载线程，当音乐播放线程被激活时，应用管理器可获取到该应用程序的包名以及音乐播放线程的名称，并发送给数据采集模块。

步骤 402：根据步骤 401 中获取到的第一应用程序的第一信息，查询数据采集配置信息，得到该第一应用程序的数据采集配置信息。

其中，数据采集配置信息中可定义需要进行数据采集的应用程序，并可进一步针对这些需要进行数据采集的应用程序，分别定义需要采集的数据（为描述方便，以下将需要采集的数据称为数据采集项）。数据采集配置信息中也可仅定义数据采集项，这种情况可适用于默认针对某一类或某些类或所有应用程序均进行数据采集。其中，数据采集项用于指定需要采集的数据，具体可通过定义采集方式或定义数据类型来指定需要采集的数据，以便以此为依据采集到需要采集的数据。一个数据采集项可以理解为用于描述被采集数据的相关属性的信息集合，根据数据采集项所采用的采集方式（比如是采集基于控件的数据还是采集网页标记性语言中的标签数据），数据采集项中可包括以下信息中的一种或多种：

基于控件的数据采集项：控件名称；

基于标签的数据采集项：标签名称。

比如，对于视频回放应用来说，如果需要采集视频播放用户界面中所显示的影片名称，且该用户界面中，影片名称由名称为“video_name”的文本框控件进行显示，则该视频回放应用对应的数据采集项中包括该文本框控件的名称“video_name”；再比如，对于新闻阅读应用来说，如果需要采集页面中所显示的新闻标题，且该页面的标记性语言源代码中（此处以 HTML 语言为例），新闻标题显示于<h1>标签内（<h1>标签用于显示新闻标题，如<h1> abc </h1>，其中 abc 表示新闻标题的文本），则该新闻阅读应用对应的数据采集项中包括<h1>标签。

需要进行数据采集的应用程序可采用应用程序列表形式，该列表中包括应用程序的识别信息，比如应用程序的包名或 URI，应用程序对应的进程名，应用程序所属类的类名，应用程序的版本信息等用于识别应用程序的信息中的一种或多种，列表中的应用程序均需要进行数据采集。数据采集配置信息中，可针对每个或每类需要进行数据采集的应用程序定义数据采集项，可采用数据采集项列表或数据采集项模板的方式进行定义，数据采集项列表或数据采集项模板中定义有需要采集的一个或多个数据项的相关配置信息。

具体实施时，可针对一类应用程序定义一个数据采集项模板，因为对于同一类应用程序来说，数据采集的需求往往具有相似性，则需要采集的数据也因而具有相同或相似的属性（比如对于在线交易类应用来说，均需采集交易额、交易时间），因此可针对同一类应用定义一个数据采集项模板，以用来针对该类应用进行数据采集。

举例来说，数据采集项中可包括以下配置信息中的一种或多种：

- 应用程序控件的名称、类型、对应的对象名称、对应的用于访问控件或对象的指令中的一种或多种。根据该配置信息，在调用数据采集接口时可通过该接口中用于获取相应控件数据的方法来得到应用程序的应用数据。
- 用于网页的标记性语言中的标签名称、标签类型中的一种或多种。根

据该配置信息，在调用数据采集接口时可通过该接口中用于获取相应网页对象中的相应标签数据的方法以获得这些标签中的应用数据。

步骤 403: 根据查询到的该第一应用程序的数据采集配置信息，基于应用程序框架层提供的接口采集该第一应用程序的应用数据。

该步骤在具体实施时，数据采集模块可首先根据步骤 401 中获取到的应用程序的包名查询应用程序列表，如果该包名在该应用程序列表中，则查询该应用程序对应的数据采集项模板，得到控件名称或标签名称，然后基于数据采集接口调用该接口中用于获取相应控件或对象中的数据的方法以获得该应用程序的应用数据。进一步地，在根据步骤 401 中获取到的应用程序的包名查询应用程序列表时，如果该包名不在该应用程序列表中，则表明无需对该应用程序进行数据采集，则放弃针对该应用程序的数据采集操作。

进一步地，在步骤 403 之后，可以将采集到的应用数据上传到服务器，也可以将采集到的应用数据保存到本地数据库中。由于采集到的应用数据中可能包含一些用户信息（比如用户账户信息、用户在线交易信息等），因此为了安全起见，可将采集到的应用数据保存到终端本地。为了更进一步提高安全性，在将应用数据存储到数据库之前，可对该应用数据进行加密。

进一步地，在步骤 403 之后，还可以根据所采集到的应用数据所归属的属性类别，对采集到的应用数据进行分类。这样，来自不同应用程序的应用数据可根据其属性归类到相应的分类中，从而基于多个应用程序的应用数据来反映用户的属性特征，一方面使得这样得到的用户特征更全面和准确，另一方面使得应用数据的存储更合理，也更便于查询和使用，进一步地，还可以将分类后的应用数据保存到终端的数据库中。其中，针对需要进行数据采集的应用程序，根据其所执行的业务所归属的类型，或者进一步结合应用数据的使用需求（比如用于用户喜好分析，还是用于对用户交易行为进行分析等），可设置一些相应的应用数据属性类别，比如，应用数据的属性类别可包括但不限于：设备属性（比如该类别的信息可包括设备或操作系统的相关信息）、用户基本属性（比如该类别的信息可包括用户的性别、年龄等信息）、

用户操作行为属性（比如该类别的信息可包括用户使用应用程序的时间、频次等）、用户喜好属性（比如该类别的信息可包括用户喜好的运动、娱乐方式等）、用户交易属性（比如该类别的信息可包括用户在线购物的商品种类、交易额等信息），在此不再一一列举。

进一步地，终端还可以接收服务器发送的数据采集配置信息，并根据接收到的数据采集配置信息更新该终端的数据采集配置信息。具体实施时，终端可接收服务器主动推送的数据采集配置信息，其中，服务器可定时或不定时推送数据采集配置信息，或者在应用程序或操作系统更新时推送数据采集配置信息。终端还可主动请求获取数据采集配置信息。具体地，终端在有应用程序被安装或有应用程序被更新时，向服务器发送针对被安装或被更新的应用程序的数据采集配置信息获取请求，并接收服务器根据该获取请求返回的数据采集配置信息。

以上仅描述了采集第一应用程序的应用数据的过程，对于运行于应用程序框架层的其他应用，比如第二应用程序，也可采用上述方式对其进行数据采集。

通过以上描述可以看出，本发明实施例中，基于应用程序框架层提供的接口，一方面可获取得到运行于操作系统应用程序框架层的多个应用程序的信息，另一方面可获得相应应用程序的应用数据，与现有技术相比，无需针对不同的应用程序制定不同的采集工具包，也无需将采集工具包集成于应用程序，从而降低了数据采集的相关处理和配置操作与应用程序的耦合度。

下面结合附图对上述架构中的各功能模块在上述流程中的作用和功能进行详细说明。

图 5 示例性地示出了数据采集模块 100 的一种结构示意图。如图所示，数据采集模块 100 中可包括如下功能单元：配置单元 102、应用信息获取单元 104、配置信息查询单元 106、数据采集单元 108、数据分类单元 110、数据存储单元 112。其中，数据分类单元 110、数据存储单元 112 为可选单元。

当数据采集模块 100 安装完成后，会触发配置单元 102 进行数据采集配

置信息的配置操作，比如应用程序列表 1021 和/或数据采集项模板 1022 的初始化操作。

当应用程序被启动或激活后，应用信息获取单元 104 被启动，通过应用程序框架层提供的接口 300 获取该应用程序的包名与当前线程名称。配置信息查询单元 106，根据应用信息获取单元 104 得到的应用程序包名与当前线程名称，在应用程序列表 1021 中进行查询，如果该应用程序存在于该应用程序列表 1021，则表明需要对该应用程序进行数据采集，否则表明不需要对该应用程序进行数据采集。若需要对该应用程序进行数据采集，则在数据采集项模板 1022 中查询该应用程序的数据采集项模板。数据采集单元 108 根据配置信息查询单元 106 查询到的数据采集项模板，基于应用程序框架层提供的接口 300（比如所提供的接口中的数据采集接口）采集相应的应用数据。数据分类单元 110，将数据采集单元 108 采集到的数据进行分类。数据存储单元 112，将分类好的数据存入数据库中。

可在操作系统安装过程中，完成数据采集的初始化工作。数据采集的初始化工作主要包括配置应用程序列表 1021 和/或数据采集项模板 1022。应用程序列表 1021 以及数据采集项模板 1022 可以预先设置在操作系统的安装包中，也可以从服务器获取，还可以在操作系统安装完成后由服务器进行更新。举例来说，当终端上安装应用程序或卸载应用程序或者有应用程序进行更新时，若该应用程序在应用程序列表 1021 中，则应用配置管理模块 200 可将数据采集配置信息推送给终端，终端中的配置单元 102 根据接收到的数据采集配置信息更新该终端的数据采集配置信息。

其中，应用程序列表 1021 中可包括需要进行数据采集的应用程序的识别信息。应用程序列表 1021 采用的数据格式可以是键值对（<KEY, VALUE>），比如，< xx 管家，com.gtgj.view>，其中，“xx 管家”为应用程序的名称，“com.gtgj.view”是该应用程序的 URI。

数据采集项模板 1022 主要用来设置应用程序中需要采集的具体内容，比如采集的内容可包括应用控件数据以及应用标签数据，均可采用 JSON 格式存

储。

应用信息获取模块 104，可通过应用程序框架层提供的接口 300，获取当前处于活动状态的进程的应用信息，包括应用程序的包名、URI、线程名、类名、版本信息等。

配置信息查询单元 106 根据应用信息获取单元 104 获取到的应用程序的包名、URI、类名、版本等信息，查询应用程序列表 1021 以判断该应用程序是否要进行数据采集。若当前应用不在应用程序列表 1021 中，则该应用不需要进行数据采集，若当前应用在应用程序列表中，则该应用需要进行数据采集。若不需要进行数据采集，则直接过滤掉，后续处理步骤不再执行。若该应用程序需要进行数据采集，则基于数据采集项模板 1022 查询该应用程序的数据采集项模板，并将该应用程序的数据采集项模板 1022 输出给数据采集单元 108。数据采集项模板 1022 中可定义两种格式的应用数据，第一种是基于操作系统提供的基本应用控件数据格式，第二种是基于操作系统提供的网页类应用标签数据。第一种数据，以 Android 操作系统为例，其对应的应用控件包括但不限于：TextView（文本视图控件）、EditText（文本编辑控件）、Button（按钮控件）、CheckBox（选择框控件）、VideoView（视频播放控件）等，该类型数据格式在系统中以对象存在，即可通过调用控件对象的 Get 方法获取应用数据。第二种数据以 webView（网页视图）为主的网页数据，具体格式可以是 HTML（Hyper Text Mark-up Language，超文本标记语言）标签组成的包含应用数据的字符串。在数据采集项模板 1022 中定义了对应上述两种数据格式的配置。

数据采集单元 108，用于筛选最终的应用数据。图 6 示例性地示出了数据采集单元 108 的数据采集过程。如图所示，在步骤 1081 中，数据采集单元 108 获得数据采集项模板 1022；在步骤 1082 中，数据采集单元 108 根据该数据采集项模板 1022 判断需要采集的数据的格式，若数据格式为应用标签数据格式，则转入步骤 1083，若数据格式为应用控件数据格式，则转入步骤 1084。在步骤 1083 中，数据采集单元 108 根据该数据采集项模板采集得到应用标签数据

格式的应用数据，具体地，对于应用标签数据格式，是以对象形式封装起来的，可根据数据采集项模板中定义的对象标识，直接调用应用程序框架层提供的接口 300 中用于该对象的方法来采集数据。在步骤 1084 中，数据采集单元根据该数据采集项模板采集得到应用控件数据格式的应用数据，具体地，对于应用控件数据格式，是以字符串展现的，可根据数据采集项模板中定义的正则表达式进行数据匹配获取数据。当然，也可能数据采集项模板中定义有多种格式的数据，比如本例子中可定义有上述两种数据格式，此种情况下，根据步骤 1083 和步骤 1084 可以得到多种数据格式的数据 1085，

数据分类单元 110 可将数据采集单元 108 采集到的应用数据进行分类处理，多个应用的数据就可以集中在一起反映用户属性特征，从而进一步为终端提供数据共享服务。表 1 示例性地示出了一种分类类别：

表 1：应用数据分类类别

序号	数据类型包括但不限于	数据内容包括但不限于
1	智能终端设备数据	手机号、运营商、系统版本、IMEI等
2	智能终端用户个人基础数据	性别、生日、婚姻状态、学历程度、月均收入、月均消费、收入明细、消费明细、星座、真实姓名、称谓、数据流量、用户邮箱、用户标签等。
3	智能终端用户家庭与工作信息	家庭住址、公司地址、公司名称等。
4	智能终端用户银行信息	所属银行客户、所属银行信用卡等。
5	智能终端用户作息時間	起床时间、睡觉时间、离家时间、到家时间、离公司时间、到公司时间。
6	智能终端用户参加的体育运动	篮球、足球、羽毛球、乒乓球、网球、游泳、登山、骑车、健身等。
7	智能终端用户喜欢的美食	自助餐、火锅、日韩料理、海鲜等。
8	智能终端用户休闲娱乐	KTV、桌游电玩、足疗按摩、温泉洗浴、养

	信息	生保健等。
9	智能终端用户喜欢的音乐	歌手、歌曲、音乐类型等。
10	智能终端用户喜欢的影视剧	影视明星、追的电视剧、观影习惯、电影类型等。
11	智能终端用户购物信息	商品种类、商品品牌、商品价格等
12	智能终端用户阅读信息	书目类型、阅读时间等。
13	智能终端用户常去地点	常去地点类型、常驻城市名称等。
14	智能终端用户出行信息	出行交通工具、入住酒店品牌、酒店等级等。
15	智能终端用户应用使用习惯	使用次数、使用时间、有时间使用规律的应用、有地点使用规律的应用等。

表 1 中，智能终端设备数据属于终端设备属性类别，智能终端用户格式基础数据属于用户基本属性类别，智能终端用户作息时间、智能终端用户喜欢的美食等属于用户喜好属性类别，智能终端用户应用使用习惯属于用户操作行为属性类别，智能终端用户购物信息属于用户交易属性类别，等等，在此不再一一举例。

数据存储单元 112 可将采集到的所有应用数据按照分类存入操作系统的加密数据库，即可以不上报服务器。

图 7 示例性地示出了服务器中的应用配置管理模块 200 的处理流程。该模块用于管理终端中的数据收集配置信息。应用配置管理模块 200 可根据应用程序列表 206，定时查询应用版本信息 202，对比应用市场 201 中的应用版本，若高于应用程序列表 206 中的版本，则进行数据采集项模板 203 的更新。

在终端侧，当用户升级应用程序的时候，判断应用程序列表 1021 中是否存在该应用程序，若存在，则比较该应用程序的版本信息。若升级版本高于应用程序列表 1021 中的版本，则请求应用配置管理模块 200 推送数据采集项模板，将该应用程序的数据采集项模板更新至数据采集项模板 1022。若用户新安装应用程序，则判断应用程序列表 1021 中是否存在该应用程序，若存在，

则通过步骤 204, 直接请求应用配置管理模块 200 推送数据采集项模板, 将该应用程序的数据采集项添加至数据采集项模板 1022 中。

在具体应用时, 在终端上, 考虑采用一种系统级数据采集方案, 采集对象是基于终端的所有第三方应用。每个应用不需单独集成采集工具包。每个应用的采集内容, 通过统一的数据采集配置信息(以配置文件形式)进行管理。对于应用的更新, 只需要更新对应的配置文件即可。采集的数据存放本地加密数据库, 不上传服务器, 从而保护用户数据的安全性。

图 8 以 Android 系统为例, 示例性地示出了一种基于应用程序框架层提供的接口进行数据采集的操作时序图。在 Android 系统中, 应用程序以 Activity 形式存在。本例子中, 应用程序框架层表示为 framework, 数据采集模块可表现为一种系统服务(本例子中称为 UserProfilingService), LbBeeService 为一种用于进行数据访问和存储的系统服务, UserProfilingService 可基于 LbBeeService 进行数据采集配置信息的获取以及对采集到的应用数据进行存储。在步骤 801 至步骤 802 中, Activity 通过调用 onResume 方法, 其状态变为运行状态(或称激活状态, 该状态下, Activity 对应的用户界面对用户可见, 且能够响应用户的操作), framework 在确定 Activity 的状态变为运行态后, 向 UserProfilingService 发送相关句柄(图中表示为 handleOnResume), 以将该 Activity 转变为运行态的信息通知给 UserProfilingService, UserProfilingService 可根据该句柄获取到该 Activity 的相关信息, 比如进程名称(或包名)、线程名称(或类名)。在步骤 803 至步骤 804 中, UserProfilingService 接收到该句柄后, 通过进程间通信, 基于 LbBeeService 获取对应的数据采集项后返回该 Activity。步骤 805 至步骤 807 中, Activity 根据该数据采集项进行数据采集, 将采集到的数据传递给 UserProfilingService, UserProfilingService 通过进程间通信, 将采集到的数据传递给 LbBeeService 进行存储。

在上述实施例的一种替换实施例中, UserProfilingService 接收到 Activity 转变为运行态的信息通知, 并获取该应用的数据采集项后, 可基于 framework 提供的接口, 根据获取到的数据采集项, 调用数据采集接口, 从而通过该接

口提供的获取方法，获取到该 Activity 的应用数据。在执行该获取方法的过程中，该 Activity 可基于该数据采集项进行数据采集，将采集到的数据传递给 UserProfilingService，UserProfilingService 通过进程间通信，将采集到的数据传递给 LbBeeService 进行存储。

为了更清楚地理解本发明的上述实施例，下面结合具体应用场景进行说明。

场景一

以采集用户喜欢的影视明星的信息为例，描述本发明实施例提供的数据采集流程。本例子中，如图 9 所示，数据采集配合信息中设置有应用程序列表[视频应用 1、视频应用 2、新闻阅读应用 3]，并针对这些应用程序分别配置有数据采集项模板。其中，视频应用 1 的数据采集项模板中定义的需要采集的内容包括：影片名称 (videoName)、影片类型 (videoType)、影片主演姓名 (videoActor)、影片播放数 (hitCount)；视频应用 2 的数据采集项模板中定义的需要采集的内容包括：影片名称 (videoName)、影片类型 (videoType)、影片主演姓名 (videoActor)、影片导演姓名 (videoDirector)、影片时长 (videoTime)；新闻阅读应用 3 的数据采集项模板中定义的需要采集的内容包括：新闻类型 (newsType)、新闻标题 (newsTitle)、新闻关键词 (keyWords)。视频应用 1 和视频应用 2 的数据采集项模板中定义的数据采集项为控件类的数据项，图 9 中示出了各数据采集项对应的控件名称，比如，视频应用 1 中的数据采集项 videoName 所对应的控件名称为 gt_account_name；新闻阅读应用 3 的数据采集项模板中定义的数据采集项为标签类的数据项，图 9 中示出了各数据采集项对应的标签名称，比如，新闻阅读应用 3 中的数据采集项 newsType 所对应的标签名称为 regular1。

本示例中，列举了用户使用视频应用 1、视频应用 2、新闻阅读应用 3 的场景，图 10 示出了上述 3 个应用程序的用户界面显示内容：

当用户使用终端中的视频应用 1 时，进入观看页面，观看名称为《极限挑战》的影片。此时该应用程序的进程处于运行状态，数据采集模块基于应

用程序框架层中的应用管理器获取视频应用 1 的包名与当前页面类名。通过应用程序列表 1021 与数据采集项模板 1022, 确定当前应用页面数据需要采集。视频应用 1 当前页面的数据与数据采集项模板中定义的数据采集指令的对应关系如表 2 所示:

表 2

应用名称	应用当前页面数据	数据采集项
视频应用1	极限挑战	gt_account_name (用于采集影片名称的指令)
	娱乐	gt_account_actor(用于采集影片类型的指令)
	孙红雷、黄渤	gt_account_actor(用于采集影片主演姓名的指令)
	1亿	gt_account_hitCount (用于采集播放次数的指令)

数据采集模块通过调用应用程序框架层提供的接口中获取控件对象数据的方法, 提取出{"videoName":"极限挑战","videoType":"娱乐","videoActor":"孙红雷、黄渤","hitCount":"100000000"}, 从而可在当前的观看页面采集到如下应用控件数据: [影片名称=极限挑战]、[影片类型=娱乐]、[影片主演姓名=孙红雷、黄渤]、[播放次数=1 亿]。数据采集模块可将上述提取的格式化数据分类存入对应的数据表中。

用户使用视频应用 2 观看了名称为《好先生》的影片, 通过本发明实施例提供的方法可在当前的观看页面采集到如下应用控件数据: [影片名称=《好先生》]、[影片类型=言情]、[影片主演姓名=孙红雷、张艺兴]、[影片导演姓名=张晓波]、[播放时长=45 分钟]。

用户使用新闻阅读应用 3 阅读了一篇娱乐新闻, 通过本发明实施例提供的方法可在当前阅读页面采集到如下应用标签数据: [新闻类型=娱乐新闻]、[新闻标题=“极限男人帮” 打响奖杯争夺 孙红雷: 最难的一次]、[关键词=孙

红雷、极限挑战]。

上述场景中，应用程序的用户界面所展示的内容与数据采集项模板中定义的采集内容可如以下表 3 所示：

表 3

应用名称	应用当前页面数据	数据采集项
视频应用1	极限挑战	videoName（影片名称）
	娱乐	videoType（影片类型）
	孙红雷、黄渤	videoActor（影片主演姓名）
	1亿	hitCount（影片播放数）
视频应用2	好先生	videoName（影片名称）
	言情	videoType（影片类型）
	孙红雷、张兴义	videoActor（影片主演姓名）
	张晓波	videoDirector（影片导演姓名）
	45分钟	videoTime（影片时长）
新闻阅读应用3	娱乐新闻	newsType（新闻类型）
	“极限男人帮”打响奖杯争夺	newsTitle（新闻标题）
	孙红雷：最难的一次	
	孙红雷、极限挑战	keyWords（新闻关键词）

场景二

终端上的第三方应用版本升级操作可能比较频繁，基于本发明实施例的数据采集方法，当应用程序版本升级时，只要将最新的数据采集配置信息（可以以配置文件形式存在）推送至终端上即可。以视频应用 1 版本更新为例，服务器中的应用配置管理模块 200 可定时去应用市场中查询视频应用 1 的版本号，如果发现应用市场中的版本号与终端上安装的版本号不一致，则对该应用进行标注，并向开发者提醒视频应用 1 需要进行数据采集配置信息的更新。开发者对比两个版本后，在配置文件中追加写入视频应用 1 数据采集项。当终端升级视频应用 1 时，对比终端中的应用程序列表 1021，若发现版本号不一致，则终端请求服务器推送数据采集配置信息，服务器将新的数据采集

配置信息推送至终端中。

图 11 示出了视频应用 1 的版本更新前和更新后的视频播放界面示意图。视频应用更新后，其视频模板界面中增加了如下显示内容：导演姓名、播放时长、评分、观众留言，相应地，数据采集项中可增加如下需要采集的数据项：导演姓名、播放时长、评分、观众留言。相应地，图 12 示出了视频应用 1 的版本更新前后的数据采集配置文件的内容，可以看出，版本更新后，配置文件中增加了新的数据采集项，表明需要采集用户界面中的导演姓名以及播放时长。

基于相同的技术构思，本发明实施例还提供了一种数据采集装置。

参见图 13，为本发明实施例提供的数据采集装置的结构示意图，该装置可实现前述实施例描述的数据采集流程。该装置可包括：应用信息获取单元 1301、配置信息查询单元 1302、数据采集单元 1303，进一步地还可包括：数据分类单元 1304、数据存储单元 1305、配置单元 1306 中的一种或多种，其中：

应用信息获取单元 1301，用于基于操作系统应用程序框架层获取第一应用程序的第一信息，以及基于操作系统应用程序框架层获取第二应用程序的第二信息；

配置信息查询单元 1302，用于根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，以及，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息；其中，所述第一信息用于识别所述第一应用程序，所述第二信息用于识别所述第二应用程序；

数据采集单元 1303，用于根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，以及，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据。

可选地，应用信息获取单元 1301 具体用于：接收应用程序框架层中的应

用管理器发送的所述第一信息，其中所述第一信息是所述应用管理器在第一应用程序激活时发送的；以及，接收应用程序框架层中的应用管理器发送的所述第二信息，其中所述第二信息是所述应用管理器在第二应用程序激活时发送的

可选地，应用管理器在第一应用程序调用重加载方法时，确定所述第一应用程序激活；所述应用管理器在第二应用程序的进程调用重加载方法时，确定所述第二应用程序激活。

可选地，第一信息包括以下信息中的一种或多种：第一应用程序的包名、第一应用程序的统一资源标识符 URI、第一应用程序对应的进程名、第一应用程序的版本；第二信息包括以下信息中的一种或多种：第二应用程序的包名、第二应用程序的 URI、第二应用程序对应的进程名、第二应用程序的版本。

可选地，所述第一信息还包括：第一应用程序的类名或线程名；所述第二信息还包括：第二应用程序的类名或线程名。

可选地，所述应用程序数据采集配置信息包括：配置为需要进行数据采集的应用程序列表；需要进行数据采集的应用程序的数据采集项配置信息，一个应用程序的数据采集项配置信息用于定义对该应用程序的哪些应用数据进行采集。

可选地，所述数据采集项配置信息包括以下信息中的一种或多种：应用程序控件的名称、类型、对应的对象名称中的一种或多种；用于网页的标记性语言中的标签名称、标签类型中的一种或多种。

可选地，数据存储单元 1305，用于在所述数据采集单元采集所述第一应用程序的应用数据之后，将采集到的第一应用程序的应用数据保存到本地数据库；以及，在所述数据采集单元采集所述第二应用程序的应用数据之后，将采集到的第二应用程序的应用数据保存到本地数据库。

可选地，数据分类单元 1304，用于在所述数据存储单元将采集到的第一应用程序的应用数据保存到本地数据库之前，根据采集到的第一应用程序的

数据所归属的属性类别，对采集到的第一应用程序的数据进行分类；以及，在所述数据存储单元将采集到的第二应用程序的应用数据保存到本地数据库之前，根据采集到的第二应用程序的数据所归属的属性类别，对采集到的第二应用程序的数据进行分类。

可选地，配置单元 1306，用于基于操作系统框架层确定有应用程序被安装或有应用程序被更新时，向服务器发送获取请求，所述获取请求用于请求获取所述被安装或所述被更新的应用程序的数据采集配置信息；以及，接收所述服务器根据所述获取请求返回的数据采集配置信息。

基于相同的技术构思，本发明实施例还提供了一种数据采集系统，该系统可包括服务器和终端。服务器用于为终端配置数据采集配置信息。所述终端用于基于操作系统应用程序框架层获取第一应用程序的第一信息，根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，其中，所述第一信息用于识别所述第一应用程序；以及，基于操作系统应用程序框架层获取第二应用程序的第二信息，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据，其中，所述第二信息用于识别所述第二应用程序。其中，所述终端和所述服务器的功能可参见前述实施例，在此不再重复。

基于相同的技术构思，本发明实施例还提供了一种数据采集装置。

参见图 14，为本发明实施例提供的数据采集装置的结构示意图，该装置可实现前述实施例描述的数据采集流程。该装置可包括：收发器 1001、处理单元 1002 和存储器 1003。

处理单元 1002 用于控制该装置的操作，包括通过收发器 1001 进行数据的传输（包括接收和/或发送）；存储器 1003 可以包括只读存储器和随机存取存储器，用于向处理单元 1002 提供指令和数据。存储器 1003 的一部分还可

以包括非易失性随机存取存储器（NVRAM）。该装置的各个组件通过总线系统耦合在一起，其中总线系统 1004 除包括数据总线之外，还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见，在图中将各种总线都标为总线系统 1004。

本申请实施例揭示的流程可以应用于处理单元 1002 中，或者由处理单元 1002 实现。在实现过程中，该装置实现的流程的各步骤可以通过处理单元 1002 中的硬件的集成逻辑电路或者软件形式的指令完成。处理单元 1002 可以是通用处理器、数字信号处理器、专用集成电路、现场可编程门阵列或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件，可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件处理器执行完成，或者用处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 1003，处理单元 1002 读取存储器 1003 中的信息，结合其硬件完成本发明实施例指示流程的步骤。

具体地，处理单元 1002 可被配置以执行上述实施例所述的数据采集流程，该流程可包括：

基于操作系统应用程序框架层获取第一应用程序的第一信息，根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，其中，所述第一信息用于识别所述第一应用程序；

基于操作系统应用程序框架层获取第二应用程序的第二信息，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据，其中，所述第二信

息用于识别所述第二应用程序。

处理单元 1002 所执行的上述流程，可参见前述实施例的描述，在此不再详述。

本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

尽管已描述了本发明的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例作出另外的变更和修改。所以，所附权利要求要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

权 利 要 求

1、一种数据采集方法，其特征在于，包括：

基于操作系统应用程序框架层获取第一应用程序的第一信息，根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，其中，所述第一信息用于识别所述第一应用程序；

基于操作系统应用程序框架层获取第二应用程序的第二信息，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据，其中，所述第二信息用于识别所述第二应用程序。

2、如权利要求 1 所述的方法，其特征在于，

基于操作系统应用程序框架层获取第一应用程序的第一信息，包括：

操作系统应用程序框架层中的应用管理器在第一应用程序激活时发送所述第一应用程序的第一信息，接收所述应用管理器发送的所述第一信息；

基于操作系统应用程序框架层获取第二应用程序的第二信息，包括：

操作系统应用程序框架层中的应用管理器在第二应用程序激活时发送所述第二应用程序的第二信息，接收所述应用管理器发送的所述第二信息。

3、如权利要求 2 所述的方法，其特征在于，所述应用管理器在第一应用程序调用重加载方法时，确定所述第一应用程序激活；

所述应用管理器在第二应用程序的进程调用重加载方法时，确定所述第二应用程序激活。

4、如权利要求 1 至 3 中任一项所述的方法，其特征在于，所述第一信息包括以下信息中的一种或多种：第一应用程序的包名、第一应用程序的统一资源标识符 URI、第一应用程序对应的进程名、第一应用程序的版本；

所述第二信息包括以下信息中的一种或多种：第二应用程序的包名、第二应用程序的 URI、第二应用程序对应的进程名、第二应用程序的版本。

5、如权利要求 4 所述的方法，其特征在于，所述第一信息还包括：第一应用程序的类名或线程名；

所述第二信息还包括：第二应用程序的类名或线程名。

6、如权利要求 1 至 5 中任一项所述的方法，其特征在于，所述应用程序数据采集配置信息包括：

配置为需要进行数据采集的应用程序列表；

需要进行数据采集的应用程序的数据采集项配置信息，一个应用程序的数据采集项配置信息用于定义对该应用程序的哪些应用数据进行采集。

7、如权利要求 6 所述的方法，其特征在于，所述数据采集项配置信息包括以下信息中的一种或多种：

应用程序控件的名称、类型、对应的对象名称中的一种或多种；

用于网页的标记性语言中的标签名称、标签类型中的一种或多种。

8、如权利要求 1 至 7 中任一项所述的方法，其特征在于，

采集所述第一应用程序的应用数据之后，还包括：将采集到的第一应用程序的应用数据保存到本地数据库；

采集所述第二应用程序的应用数据之后，还包括：将采集到的第二应用程序的应用数据保存到本地数据库。

9、如权利要求 8 所述的方法，其特征在于，

将采集到的第一应用程序的应用数据保存到本地数据库之前，还包括：根据采集到的第一应用程序的数据所归属的属性类别，对采集到的第一应用程序的数据进行分类；

将采集到的第二应用程序的应用数据保存到本地数据库之前，还包括：根据采集到的第二应用程序的数据所归属的属性类别，对采集到的第二应用程序的数据进行分类。

10、如权利要求 1 至 9 中任一项所述的方法，其特征在于，还包括：

基于操作系统框架层确定有应用程序被安装或有应用程序被更新时，向服务器发送获取请求，所述获取请求用于请求获取所述被安装或所述被更新的应用程序的数据采集配置信息；

接收所述服务器根据所述获取请求返回的数据采集配置信息。

11、一种数据采集装置，其特征在于，包括：

应用信息获取单元，用于基于操作系统应用程序框架层获取第一应用程序的第一信息，以及基于操作系统应用程序框架层获取第二应用程序的第二信息；

查询单元，用于根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，以及，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息；其中，所述第一信息用于识别所述第一应用程序，所述第二信息用于识别所述第二应用程序；

数据采集单元，用于根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，以及，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据。

12、如权利要求 11 所述的装置，其特征在于，所述应用信息获取单元具体用于：接收应用程序框架层中的应用管理器发送的所述第一信息，其中所述第一信息是所述应用管理器在第一应用程序激活时发送的；以及，接收应用程序框架层中的应用管理器发送的所述第二信息，其中所述第二信息是所述应用管理器在第二应用程序激活时发送的。

13、如权利要求 12 所述的装置，其特征在于，所述应用管理器在第一应用程序调用重加载方法时，确定所述第一应用程序激活；

所述应用管理器在第二应用程序的进程调用重加载方法时，确定所述第二应用程序激活。

14、如权利要求 11 至 13 中任一项所述的装置，其特征在于，所述第一

信息包括以下信息中的一种或多种：第一应用程序的包名、第一应用程序的统一资源标识符 URI、第一应用程序对应的进程名、第一应用程序的版本；

所述第二信息包括以下信息中的一种或多种：第二应用程序的包名、第二应用程序的 URI、第二应用程序对应的进程名、第二应用程序的版本。

15、如权利要求 14 所述的装置，其特征在于，所述第一信息还包括：第一应用程序的类名或线程名；

所述第二信息还包括：第二应用程序的类名或线程名。

16、如权利要求 11 至 15 中任一项所述的装置，其特征在于，所述应用程序数据采集配置信息包括：

配置为需要进行数据采集的应用程序列表；

需要进行数据采集的应用程序的数据采集项配置信息，一个应用程序的数据采集项配置信息用于定义对该应用程序的哪些应用数据进行采集。

17、如权利要求 16 所述的装置，其特征在于，所述数据采集项配置信息包括以下信息中的一种或多种：

应用程序控件的名称、类型、对应的对象名称中的一种或多种；

用于网页的标记性语言中的标签名称、标签类型中的一种或多种。

18、如权利要求 11 至 17 中任一项所述的装置，其特征在于，还包括：

数据存储单元，用于在所述数据采集单元采集所述第一应用程序的应用数据之后，将采集到的第一应用程序的应用数据保存到本地数据库；以及，在所述数据采集单元采集所述第二应用程序的应用数据之后，将采集到的第二应用程序的应用数据保存到本地数据库。

19、如权利要求 18 所述的装置，其特征在于，还包括：

数据分类单元，用于在所述数据存储单元将采集到的第一应用程序的应用数据保存到本地数据库之前，根据采集到的第一应用程序的数据所归属的属性类别，对采集到的第一应用程序的数据进行分类；以及，在所述数据存储单元将采集到的第二应用程序的应用数据保存到本地数据库之前，根据采集到的第二应用程序的数据所归属的属性类别，对采集到的第二应用程序的

数据进行分类。

20、如权利要求 11 至 19 中任一项所述的装置，其特征在于，还包括：

配置单元，用于基于操作系统框架层确定有应用程序被安装或有应用程序被更新时，向服务器发送获取请求，所述获取请求用于请求获取所述被安装或所述被更新的应用程序的数据采集配置信息；以及，接收所述服务器根据所述获取请求返回的数据采集配置信息。

21、一种数据采集系统，其特征在于，包括：

服务器，用于为终端配置数据采集配置信息；

终端，用于基于操作系统应用程序框架层获取第一应用程序的第一信息，根据所述第一信息查询应用程序数据采集配置信息，得到第一应用程序的数据采集配置信息，根据所述第一应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第一应用程序的应用数据，其中，所述第一信息用于识别所述第一应用程序；以及，基于操作系统应用程序框架层获取第二应用程序的第二信息，根据所述第二信息查询应用程序数据采集配置信息，得到第二应用程序的数据采集配置信息，根据所述第二应用程序的数据采集配置信息并基于操作系统应用程序框架层提供的接口采集所述第二应用程序的应用数据，其中，所述第二信息用于识别所述第二应用程序。

22、一种通信装置，其特征在于，包括：处理单元和存储器；

所述存储器，用于存储计算机程序指令；

所述处理单元，耦合到所述存储器，用于读取所述存储器存储的计算机程序指令，并执行如权利要求 1 至 10 中任意一项所述的方法。

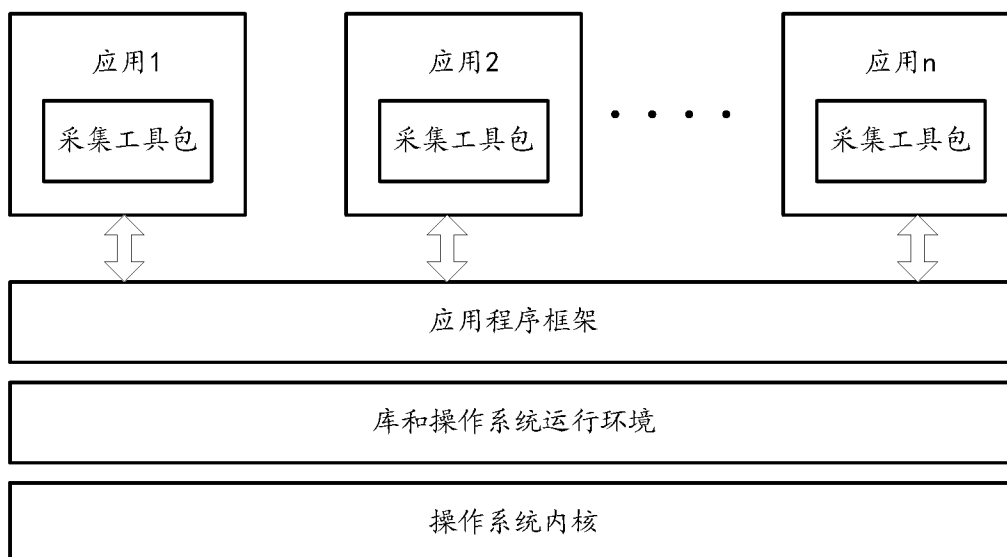


图 1

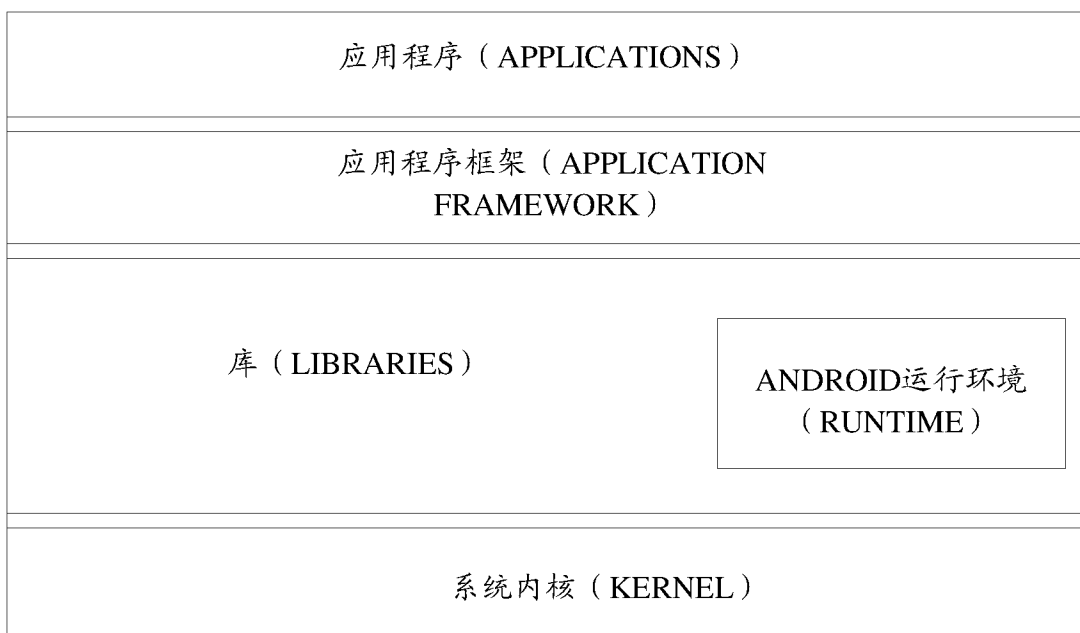


图 2

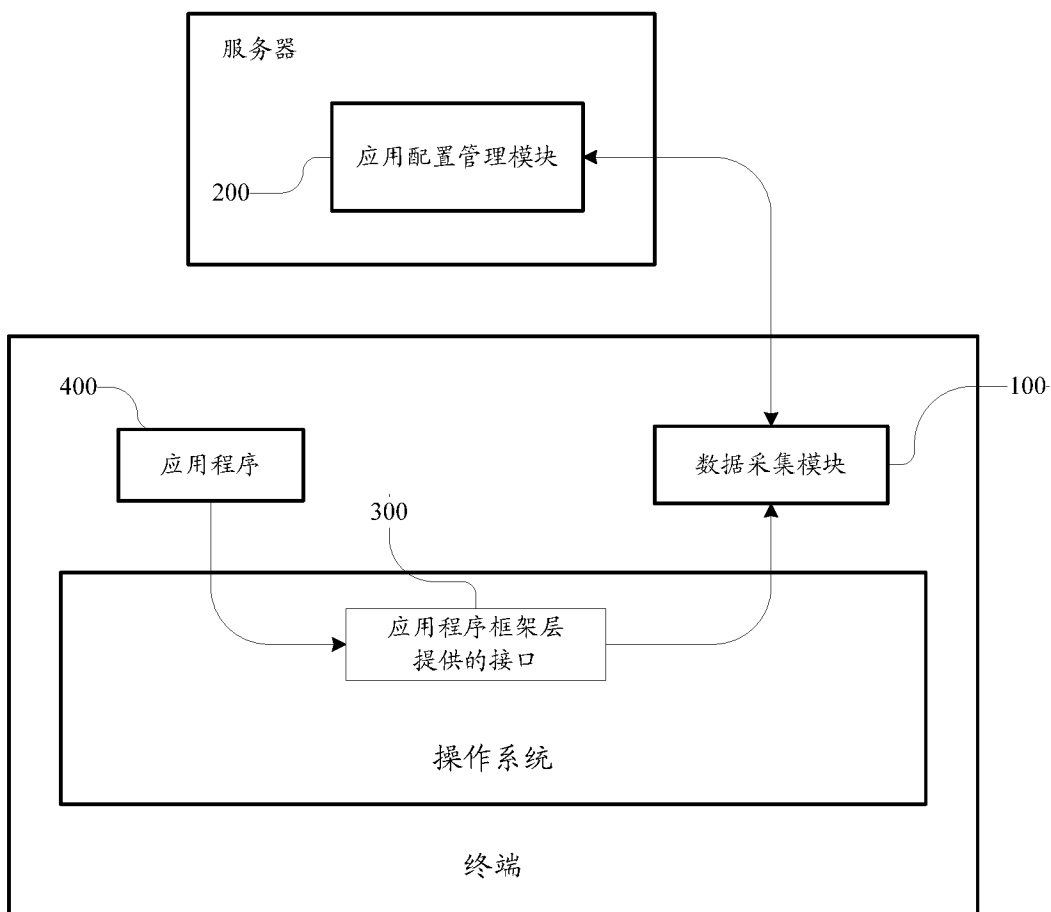


图 3

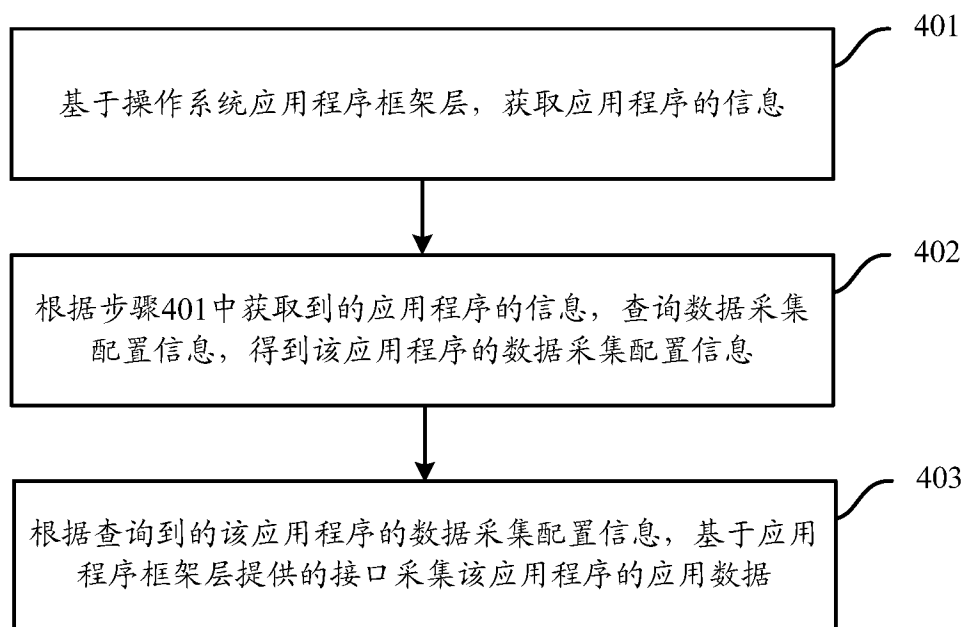


图 4

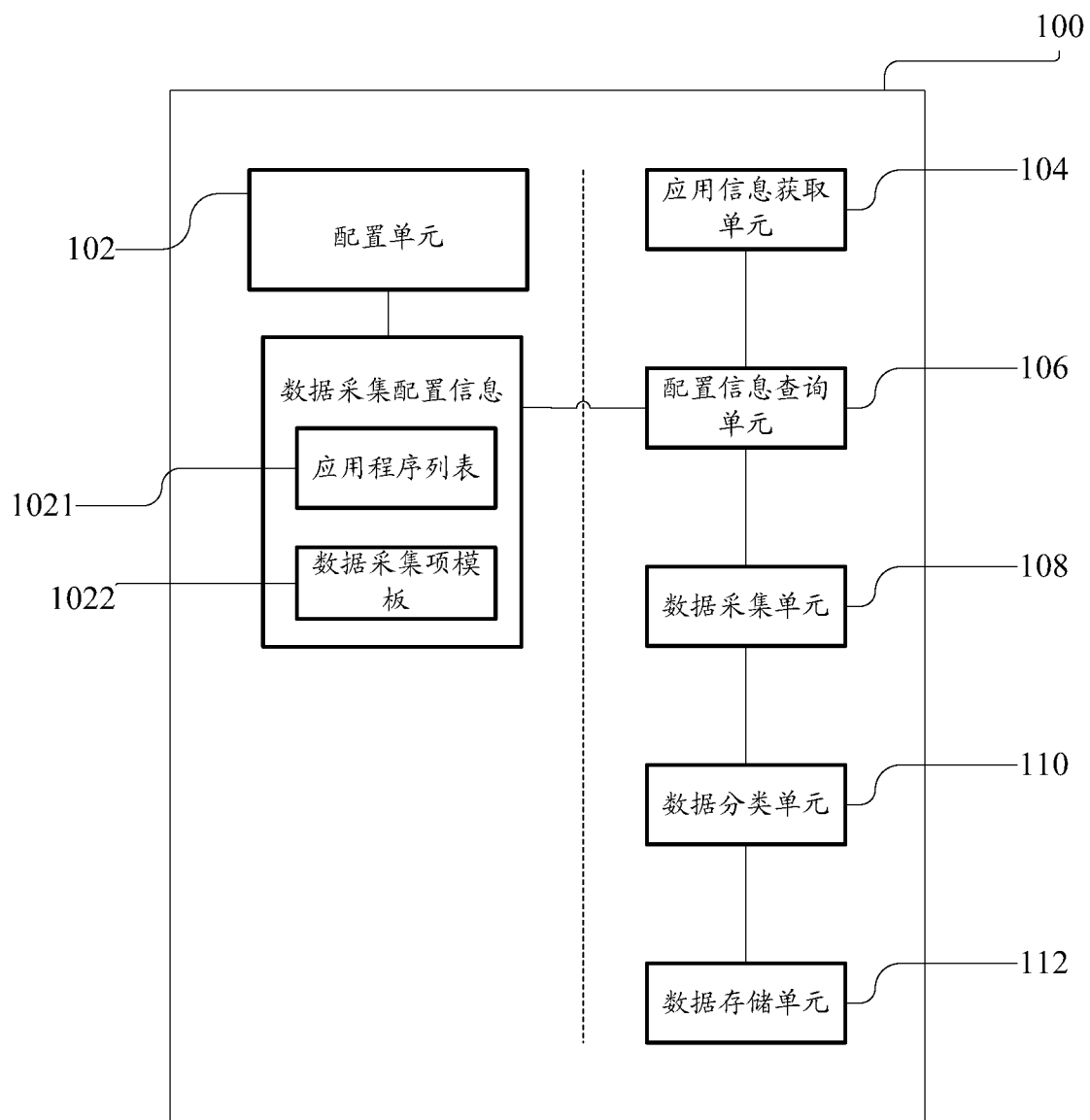


图 5

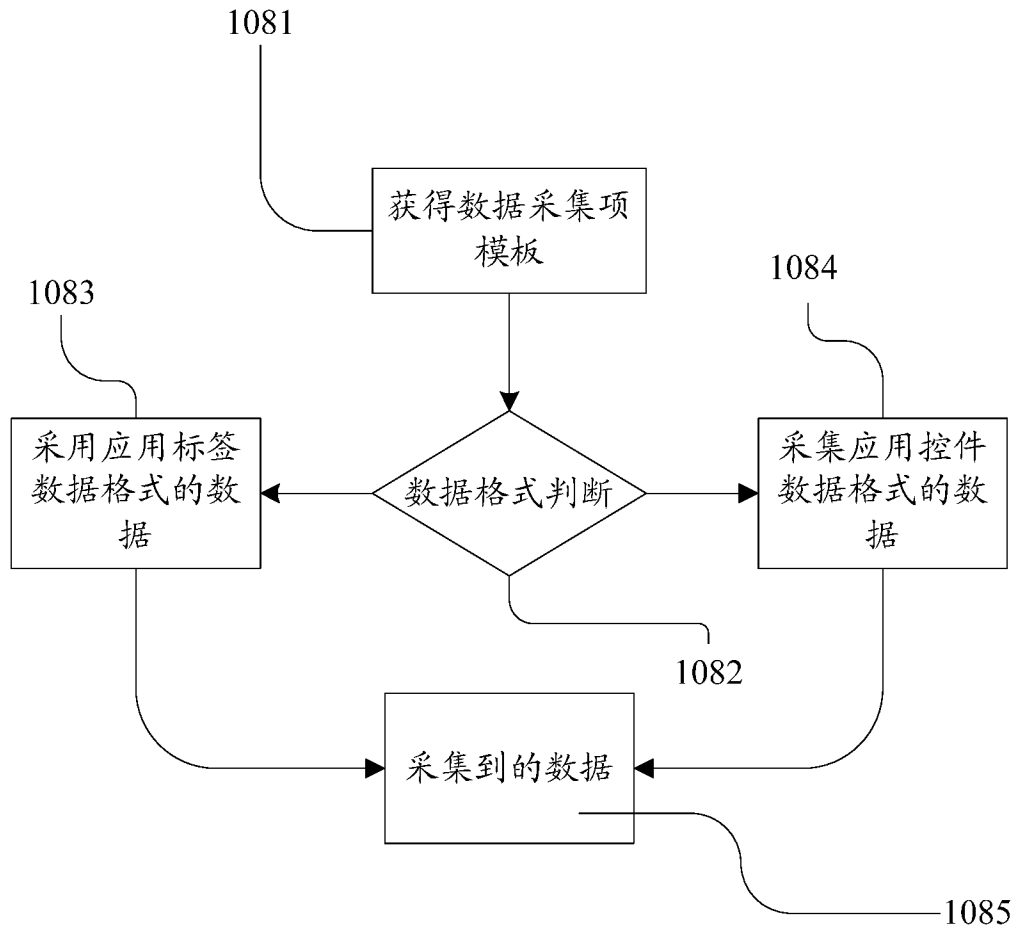


图 6

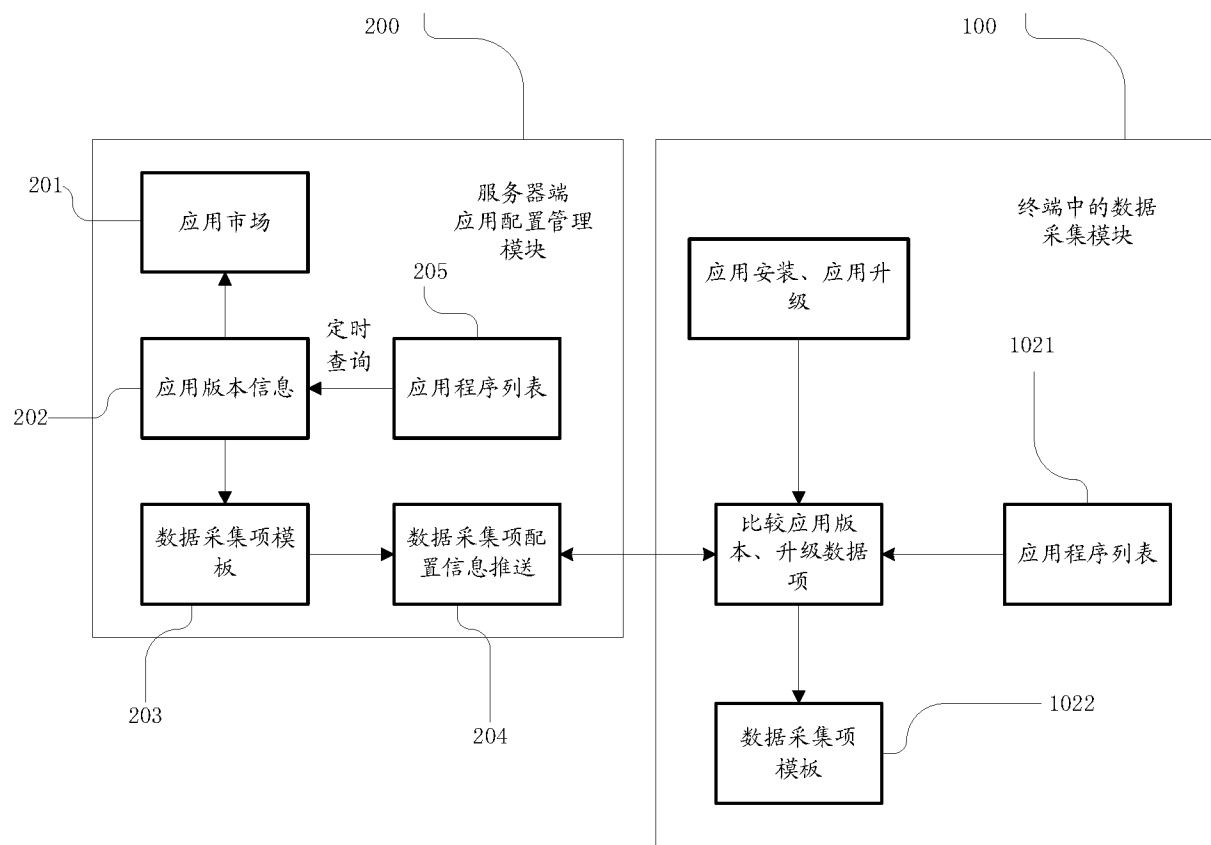


图 7

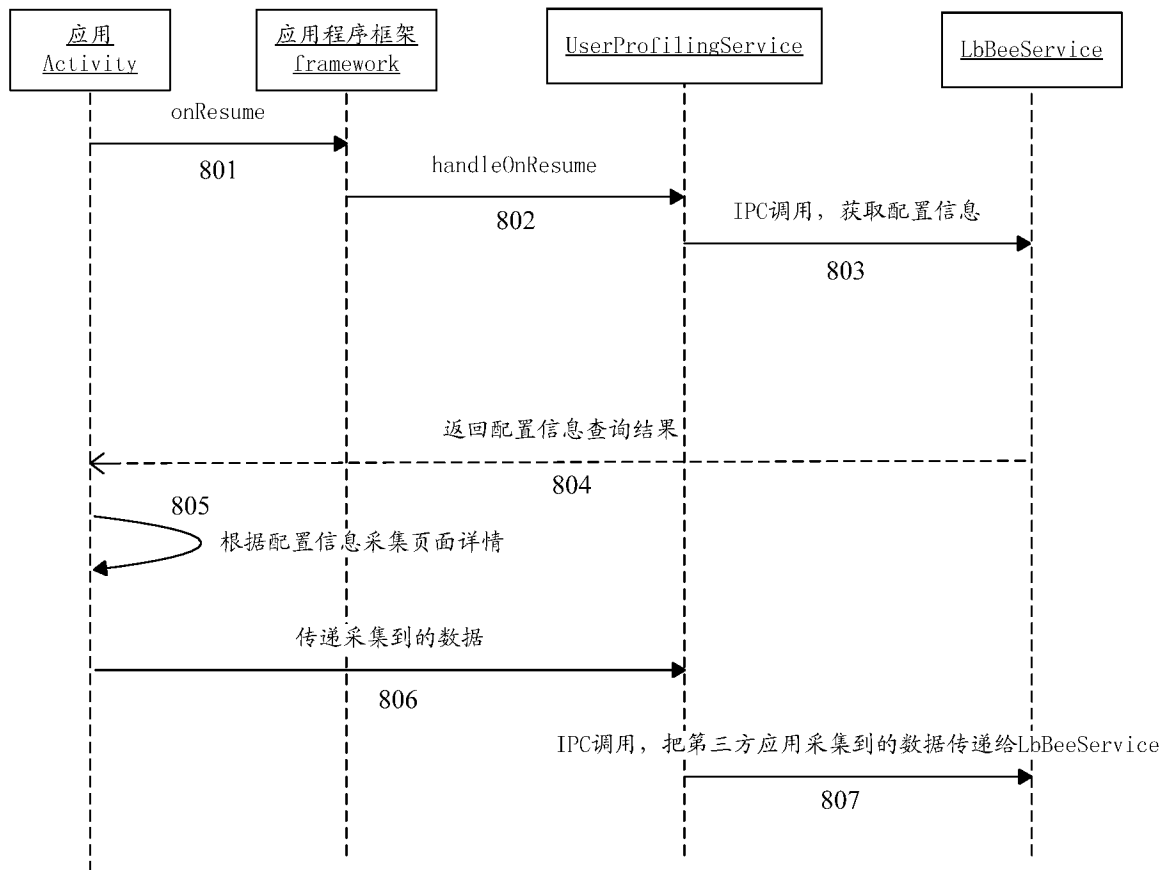


图 8

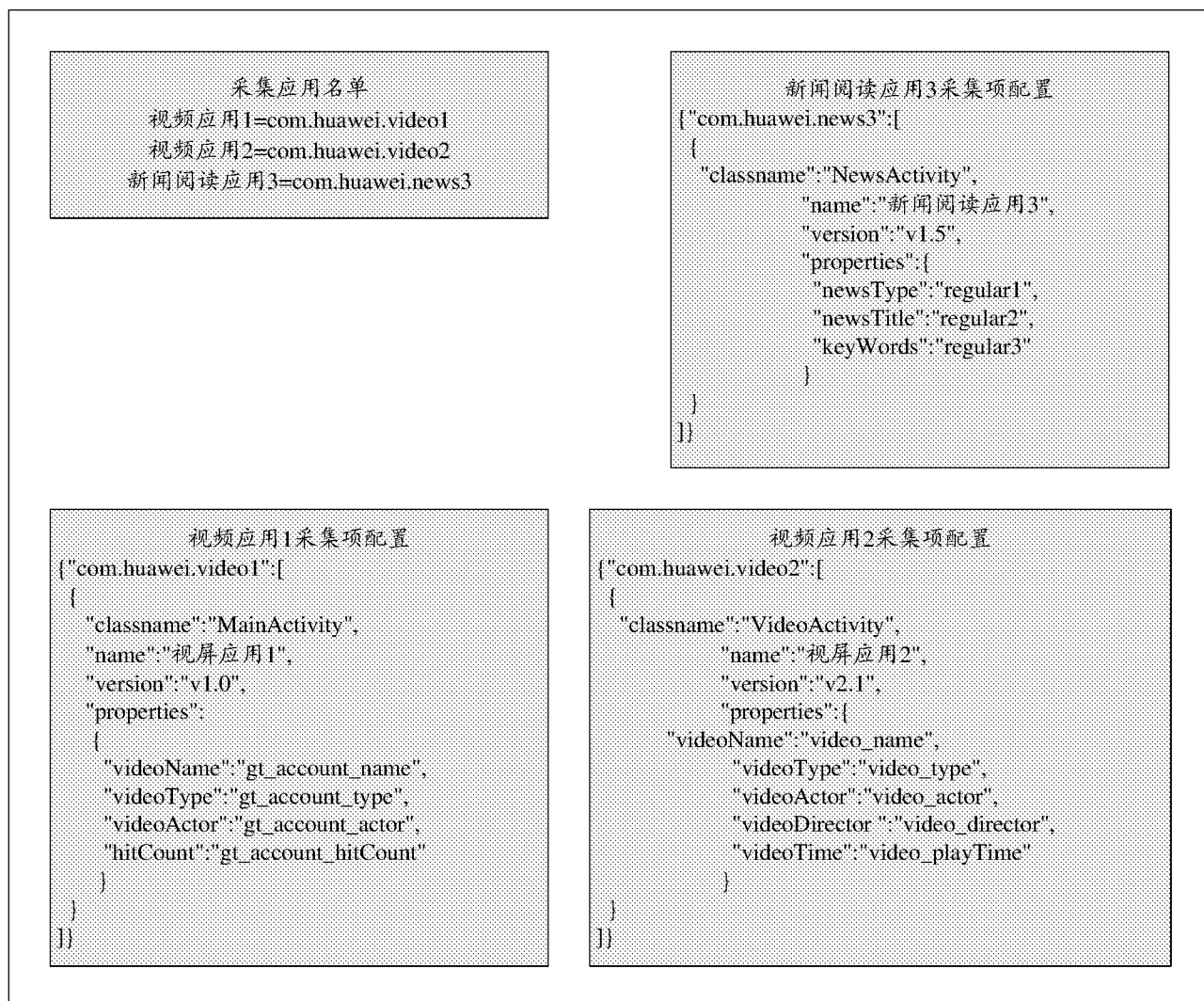


图 9



图 10

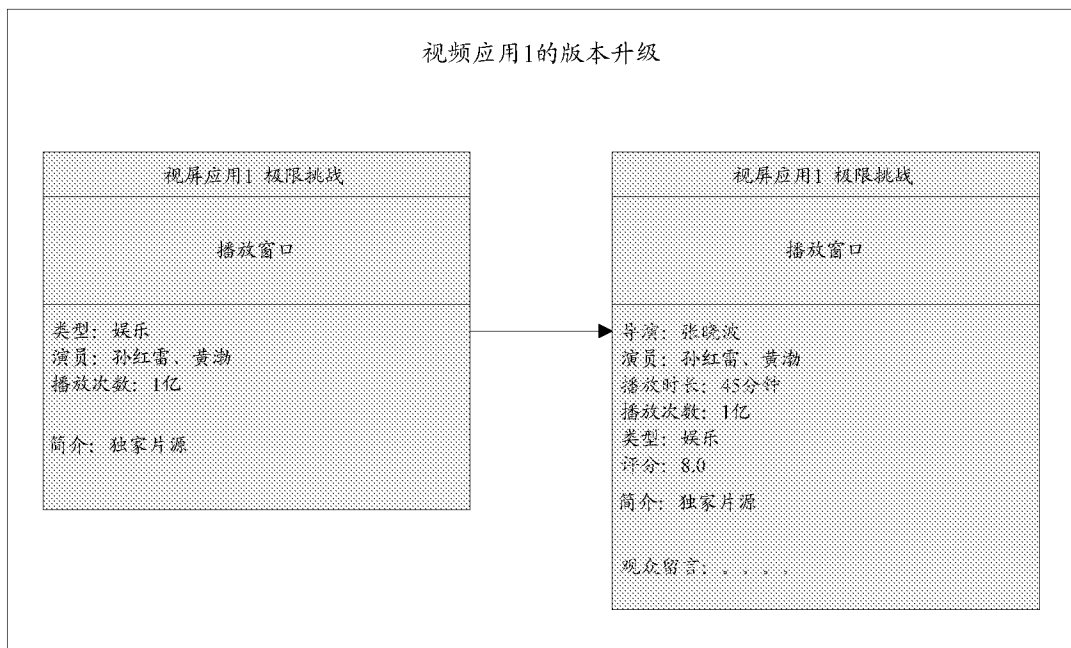


图 11

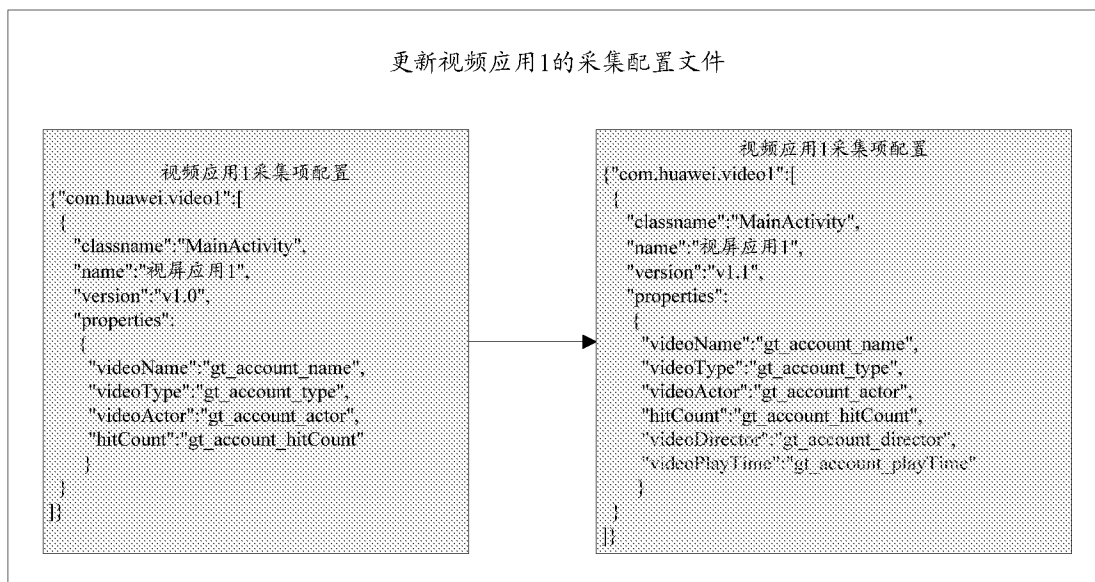


图 12

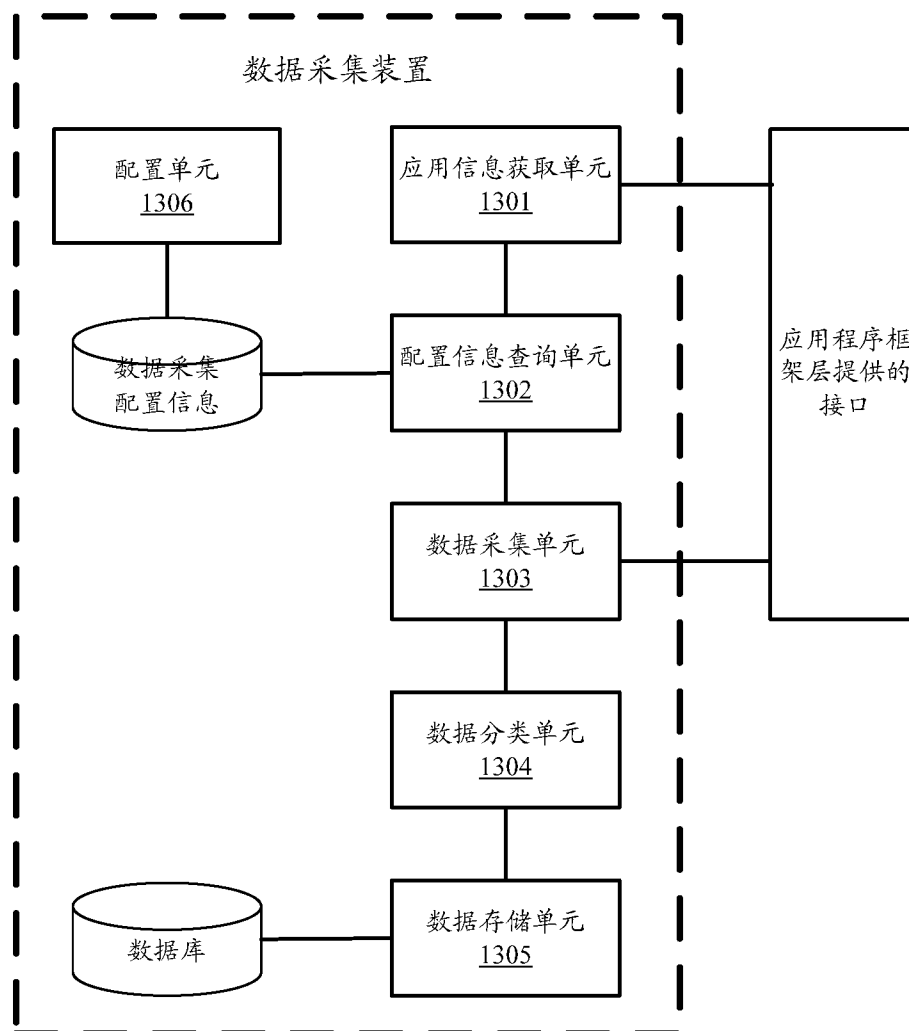


图 13

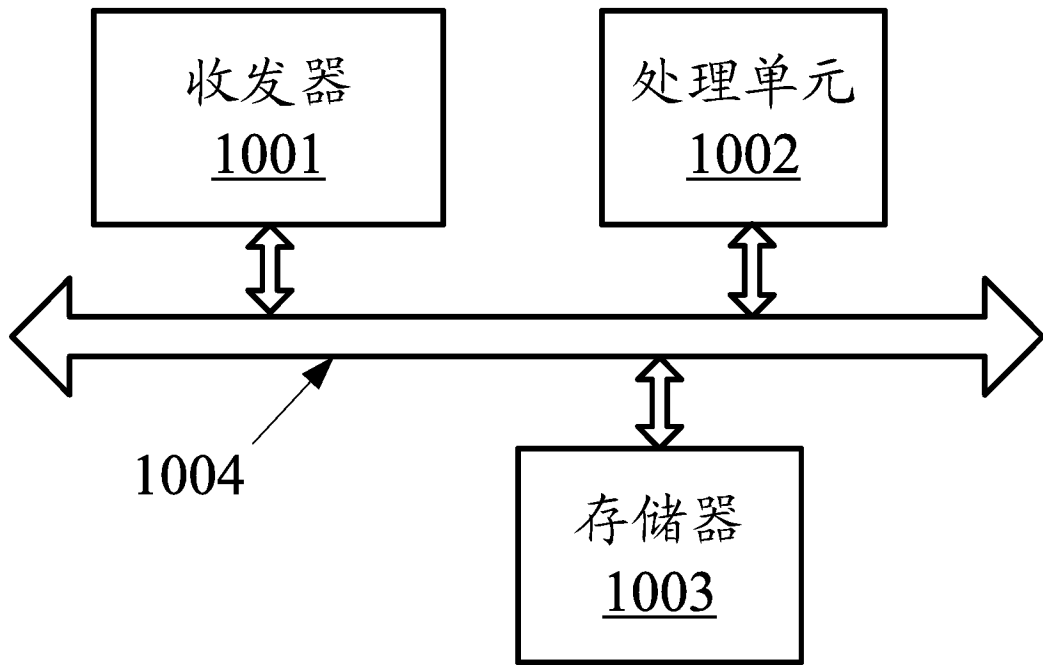


图 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/098240

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/44 (2006.01) i; G06F 17/40 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: 应用, 程序, 接口, 数据, 采集, 框架, 包名, 统一资源标识符, 进程名, 版本, application, program, API, APP, interface, collect, gather, data, frame, framework, packet, name, URI, version

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103942054 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 23 July 2014 (23.07.2014), description, paragraphs [0037]-[0046]	1-22
A	CN 103455520 A (BEIJING SAMSUNG COMMUNICATION TECHNOLOGY RESEARCH CO., LTD. et al.) 18 December 2013 (18.12.2013), entire document	1-22
A	CN 102314348 A (TCL CORPORATION) 11 January 2012 (11.01.2012), entire document	1-22
A	WO 2013140076 A2 (ABLE FRANCE) 26 September 2013 (26.09.2013), entire document	1-22

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 02 May 2017	Date of mailing of the international search report 05 June 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer WANG, Jian Telephone No. (86-10) 61648279

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/098240

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103942054 A	23 July 2014	None	
CN 103455520 A	18 December 2013	None	
CN 102314348 A	11 January 2012	None	
WO 2013140076 A2	26 September 2013	US 2015242538 A1	27 August 2015
		WO 2013140077 A1	26 September 2013
		CN 104321768 A	28 January 2015
		CN 104321743 A	28 January 2015
		US 2015161219 A1	11 June 2015
		FR 2988192 A1	20 September 2013

国际检索报告

国际申请号

PCT/CN2016/098240

A. 主题的分类

G06F 9/44(2006.01)i; G06F 17/40(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, WPI, EPODOC, CNKI: 应用, 程序, 接口, 数据, 采集, 框架, 包名, 统一资源标识符, 进程名, 版本, application, program, API, APP, interface, collect, gather, data, frame, framework, packet, name, URI, version

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 103942054 A (北京邮电大学) 2014年 7月 23日 (2014 - 07 - 23) 说明书第[0037]-[0046]段	1-22
A	CN 103455520 A (北京三星通信技术研究有限公司 等) 2013年 12月 18日 (2013 - 12 - 18) 全文	1-22
A	CN 102314348 A (TCL集团股份有限公司) 2012年 1月 11日 (2012 - 01 - 11) 全文	1-22
A	WO 2013140076 A2 (ABLE FRANCE) 2013年 9月 26日 (2013 - 09 - 26) 全文	1-22

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2017年 5月 2日

国际检索报告邮寄日期

2017年 6月 5日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

王健

电话号码 (86-10)61648279

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/098240

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103942054	A	2014年 7月 23日	无			
CN	103455520	A	2013年 12月 18日	无			
CN	102314348	A	2012年 1月 11日	无			
WO	2013140076	A2	2013年 9月 26日	US	2015242538	A1	2015年 8月 27日
				WO	2013140077	A1	2013年 9月 26日
				CN	104321768	A	2015年 1月 28日
				CN	104321743	A	2015年 1月 28日
				US	2015161219	A1	2015年 6月 11日
				FR	2988192	A1	2013年 9月 20日

表 PCT/ISA/210 (同族专利附件) (2009年7月)