



(12) 发明专利

(10) 授权公告号 CN 101405696 B

(45) 授权公告日 2012. 03. 21

(21) 申请号 200780010137. 3

(22) 申请日 2007. 02. 01

(30) 优先权数据

06112793. 2 2006. 04. 20 EP

(85) PCT申请进入国家阶段日

2008. 09. 22

(86) PCT申请的申请数据

PCT/EP2007/050974 2007. 02. 01

(87) PCT申请的公布数据

W02007/122021 EN 2007. 11. 01

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 R·甘杰米 L·皮克逊 A·泰伦兹

A·卡波马西 M·维塔莱蒂

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 于静 李峥

(56) 对比文件

US 6366930 B1, 2002. 04. 02, 全文.

WO 0205184 A2, 2002. 01. 17, 全文.

WO 0153938 A1, 2001. 07. 26, 文件说明书第 3-11 页.

审查员 徐卫锋

(51) Int. Cl.

G06F 9/445 (2006. 01)

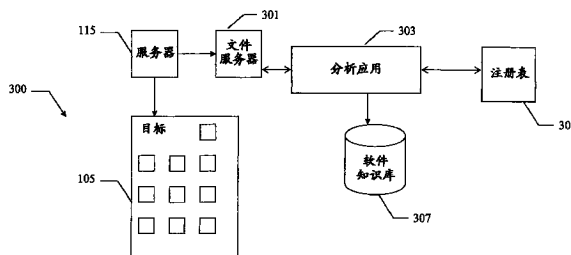
权利要求书 2 页 说明书 5 页 附图 4 页

(54) 发明名称

用于生成产品签名的方法和系统

(57) 摘要

提出了一种用于在多个计算机上发现软件产品的方法和对应的装置。所述发现系统的目的是在没有管理员的手动干预的情况下填充软件目录。本发明减少了产生和维护全面知识库(目录)的成本,所述知识库包含软件产品的定义和相关签名。签名定义结合了从数据收集过程获得的信息以及在注册表中包含的信息。此结合允许健壮签名的创建,所述健壮签名很大程度地降低了“假阳性”和“假阴性”结果的可能性。



1. 一种用于生成产品签名的方法,所述产品签名指示出相关联软件产品在多个数据处理系统上的存在,所述方法包括以下步骤:

扫描所述多个数据处理系统的文件系统,以检测核心文件集合的出现,所述集合的每个指示出被安装在数据处理系统上的产品;

对于所述集合的每个,收集注册表信息的至少一个单元;

确定关联于软件产品的对应产品签名,所述签名包括:指示出所述核心文件集合的基于文件部分和指示出注册表信息的至少一个单元的基于注册表部分,

其中所述扫描所述多个数据处理系统的文件系统,以检测核心文件集合的出现包括以下步骤:

在所述多个数据处理系统上标识多个可执行文件;

创建包含所述多个已标识可执行文件的列表;

在所述列表上选择至少一个文件;

对于每个所选择文件,创建足迹集合,所述足迹集合包括在包含所选择文件的所有文件夹中出现的文件。

2. 根据权利要求1所述的方法,其中所述注册表信息的至少一个单元从对应的文件集合的至少一个文件中得出。

3. 根据权利要求1所述的方法,进一步包括以下步骤:

根据每个文件的出现数量对所述列表排序;

从具有最高出现的文件开始,在所述列表上选择一个文件,直到所述列表被扫描完,并且对于每个所选择的文件:

创建足迹集合,所述足迹集合包括在包含所选择文件的所有文件夹中出现的文件;

从所述已排序列表中删除被包括在所述足迹中的所有文件。

4. 根据权利要求1-3任一项所述的方法,其中所述产品签名的基于文件部分包括:所述文件集合的文件名称和相对于公共父文件夹的路径。

5. 根据权利要求1-3任一项所述的方法,其中收集注册表信息的至少一个单元的所述步骤包括:根据对应的文件集合的至少一个文件查询操作系统注册表。

6. 根据权利要求1-3任一项所述的方法,其中所述签名与软件产品标识符的关联是基于由在对应的文件集合中包括的文件所提取的信息。

7. 一种用于生成产品签名的系统,所述产品签名指示出相关联软件产品在多个数据处理系统上的存在,所述系统包括:

用于扫描所述多个数据处理系统的文件系统,以检测核心文件集合的出现的装置,所述集合的每个指示出被安装在数据处理系统上的产品;

用于对于所述集合的每个,收集注册表信息的至少一个单元的装置;

用于确定关联于软件产品的对应产品签名的装置,所述签名包括:指示出所述核心文件集合的基于文件部分和指示出注册表信息的至少一个单元的基于注册表部分,

其中用于所述扫描所述多个数据处理系统的文件系统,以检测核心文件集合的出现的装置包括:

用于在所述多个数据处理系统上标识多个可执行文件的部件;

用于创建包含所述多个已标识可执行文件的列表的部件;

用于在所述列表上选择至少一个文件的部件；

用于对于每个所选择文件，创建足迹集合的部件，所述足迹集合包括在包含所选择文件的所有文件夹中出现的文件。

8. 根据权利要求 7 所述的系统，其中所述注册表信息的至少一个单元从对应的文件集合的至少一个文件中得出。

9. 根据权利要求 7 所述的系统，进一步包括

用于根据每个文件的出现数量对所述列表排序的装置；

用于从具有最高出现的文件开始，在所述列表上选择一个文件，直到所述列表被扫描完，并且对于每个所选择的文件：

创建足迹集合，所述足迹集合包括在包含所选择文件的所有文件夹中出现的文件；

从所述已排序列表中删除被包括在所述足迹中的所有文件的装置。

10. 根据权利要求 7-9 任一项所述的系统，其中所述产品签名的基于文件部分包括：所述文件集合的文件名称和相对于公共父文件夹的路径。

11. 根据权利要求 7-9 任一项所述的系统，其中所述用于对于所述集合的每个，收集注册表信息的至少一个单元的装置包括：用于根据对应的文件集合的至少一个文件查询操作系统注册表的部件。

12. 根据权利要求 7-9 任一项所述的系统，其中所述签名与软件产品标识符的关联是基于由在对应的文件集合中包括的文件所提取的信息。

用于生成产品签名的方法和系统

技术领域

[0001] 本发明涉及数据处理领域。更具体地,本发明涉及一种用于在分布式数据处理系统上检测软件产品的存在的方法。具体而言,由本发明解决了签名生成的行为。本发明进一步涉及一种用于执行所述方法的计算机程序,以及一种包含所述程序的产品。而且,本发明还涉及一种对应的装置。

[0002] 背景技术

[0003] 一些系统管理软件应用需要包含关于软件产品的信息的全面知识库的可用性,所述软件产品可以当前安装并运行在多个数据处理系统上。例如,像 IBM Tivoli License Manager (ITLM) 的许可管理器产品需要知识库(即目录)来标识在所管理系统上发现的产品,以及正确地计量这些产品的使用并为其开账单。通常,该知识库采用包含软件产品的定义(例如产品名称和版本)以及相关签名的目录的形式。软件目录列出了可在所管理系统上找到的所有已知产品;每个产品可以由指示其正在运行的一个或多个可执行模块来标识。作为示例,在许可管理器产品的情形中,在后台工作的许可代理检测已启动的可执行模块;许可代理接着通过软件目录来标识对应产品。

[0004] 现今大量使用的一种用于获得安装在计算机系统上的软件的清单(inventory)的方法是运行一种清点应用,其尝试用预定义的签名与扫描某些已知注册表、文件系统或二者的结合的结果进行匹配。创建将在上述过程中使用的签名通常需要对将发现的产品的深入知识。实际上,依赖于注册表签名可能导致假阳性,例如,当软件产品的卸载在注册表中留下孤立信息的时候。指定关键可执行产品的名称和大小的文件签名将极少产生假阳性。然而,因为下述因素,可以容易地生成假阴性,所述因素为:可以改变文件大小(更新)或使其不可预测(文件由静态链接的首要库所建立,所述静态链接的首要库取决于其版本具有不同大小)。

[0005] 指定关键文件的大小(或校验和)和名称的产品签名不会产生假阳性,但是存在用于根据由每次新产品升级和补丁所引入的、对于该文件的大小或校验和的改变而保持产品签名目录(或知识库)为最新的很高成本。在实践中,针对已经在领域中部署的签名文件的每个变体,在目录中必须存在一个签名。

[0006] 本发明的目的是提供一种减轻上述缺陷的方法和系统。

发明内容

[0007] 根据本发明,我们提供了一种用于生成产品签名的方法,所述产品签名指示出相关联软件产品在多个数据处理系统上的存在,所述方法包括以下步骤:扫描所述多个数据处理系统的文件系统,以检测文件集合的出现,所述集合的每个指示出被安装在数据处理系统上的产品;对于所述集合的每个,收集注册表信息的至少一个单元;确定关联于软件产品的对应产品签名,所述签名包括:指示出所述文件集合的基于文件部分和指示出注册表信息的至少一个单元的基于注册表部分。

[0008] 本发明的另一方面提供了一种用于跨多个数据处理系统发现软件产品的方法。

[0009] 本发明的再一方面提供了一种用于执行上述方法的计算机程序。

[0010] 本发明的又一方面提供了一种包含该程序的程序产品。

[0011] 此外,本发明的另一方面提供了一种用于实现上述方法的对应装置。

[0012] 根据本发明的优选实施例,提供了一种用于针对软件产品指定签名的方法,所述签名不需要被更新以反映由针对该产品而发布的软件更新和修复所导致的改变。关键在于利用某些相关文件的存在(仅由其名称和其相对于公共安装文件夹的相对路径所标识)作为产品实例被安装在所扫描系统上的指示符。此信息当独立使用时不足以确定所检测实例的版本,因为具有相同名称的文件可能在产品的不同发布(release)中为公共的。因此,实例的版本从关联于该实例的注册表数据的值得出。在本发明的优选实施例中,签名的生成需要总是产品实例的一部分的核心文件集合的标识,而不管在安装期间指定的选项。此标识基于对从其中安装了软件产品的计算机的典型样本收获(harvest)的信息的自动化分析。根据本发明,单一签名可以检测对应于不同版本或发布的相同产品。因此,所提出的类型的一个签名在软件知识库中通常关联于若干产品-版本定义。在本发明的优选实施例中,对应于所检测的特定实例的产品版本从所指定的注册表键获得,所述注册表键的值需要在所扫描的计算机上获得。公共安装文件夹的绝对路径名(位置)通常也从关联于同一产品实例的注册表数据获得,相对于所述公共安装文件夹,签名文件的所有相对路径被指定。在此情形中,搜索匹配实例的过程将是“即时的”,因为将不需要对文件系统的完全扫描,而是仅需要对于具有指定名称和相对路径的文件集合实际上在该位置下确实存在的验证。某些产品可能不需要将其版本和安装路径信息注册到平台软件注册表上,以便成为可运行的。当通过已提出的签名搜索非已注册软件实例时,基于来自对文件系统的完全扫描的数据,签名规范的基于文件部分仍可以匹配,即,仅仅签名的“即时性”特征丢失。在此情形中,签名将不会返回所检测实例的版本的值,于是仅可以基于软件知识库中的相关联产品-版本定义来确定针对此属性的可能值的范围。

[0013] 在所附权利要求中阐述了被确认是新颖特征的本发明的特性。然而,通过参考将结合附图阅读的以下详细描述,将最佳地理解本发明本身及其这些和其它相关目的和优点。

附图说明

[0014] 图 1 是其中可应用本发明的方法的数据处理系统的示意性框图;

[0015] 图 2 示出了系统的通用计算机的功能块;

[0016] 图 3 描述了可以用于实践方法的主要组件;以及

[0017] 图 4 示出了描述与方法的示意性实现相关的动作流的图。

具体实施方式

[0018] 具体参考图 1,描述了实现软件清单应用的、具有分布式体系结构的数据处理系统 100。系统 100 包括多个端点 105,其被分组为多个集合。端点 105 可以是工作站、服务器、膝上型计算机、以及更广泛地是其中安装了多个软件产品的任意类型的数据处理系统。端点 105 通过网络 110(例如 LAN)彼此连接。端点 105 的不同集合通过网络 120(例如基于因特网)与系统管理服务器 115 通信;系统管理服务器 115 实现了中央储存库,其中系统

100 的清单数据被收集和管理。如上所述,本发明可以在许可管理系统中实现,但更广泛地可以在需要清点动作的任意系统中实现,所述系统例如,需要在已清点了每个工作站的当前状态之后将修复或增强分布到多个工作站的软件部署系统;另一种可能的使用是在配置管理系统中。

[0019] 如图 2 所示,通过 150 示出了系统的通用计算机(工作站、本地服务器、或者系统管理服务器)。计算机 150 由并行连接于系统总线 153 的若干单元组成。详细地,一个或多个微处理器(μP) 156 控制计算机 150 的操作;RAM 159 被微处理器 156 直接用作工作存储器,以及 ROM 162 存储用于计算机 150 的自展的基本代码。外围单元在本地总线 165 附近聚集(通过各个接口)。具体地,海量存储器包括硬盘 168 和用于读取 CD-ROM 174 的驱动器 171。此外,计算机 150 包括输入设备 177(例如键盘和鼠标)以及输出设备 180(例如监视器和打印机)。网络接口卡(NIC) 183 用于将计算机 150 连接到网络。桥单元 186 将系统总线 153 与本地总线 165 接合。每个微处理器 156 和桥单元 186 可以作为主代理运行,所述主代理请求对系统总线 153 的访问以传送信息。判优器 189 管理对于系统总线 153 的互斥访问的授权。

[0020] 如果系统具有不同拓扑或者其基于其它网络,类似的考虑也可以应用。可替换地,计算机具有不同结构,包括等价单元,或者包括其它数据处理实体(诸如 PDA、移动电话等)。在任意情形中,本发明的方案同样适用于其中对工作站的控制被分散的系统中,或者甚至适用于单独计算机中。

[0021] 现在考虑图 3,通过 300 总体上描述了可用于实施本发明的优选实施例的方法的主要组件。信息(程序和数据)典型被存储在不同计算机的硬盘上,并当程序正在运行时被(至少部分地)加载到相应的工作存储器中。

[0022] 系统管理服务器 115 控制和执行对通信网络所连接的若干端点 105 的管理操作。端点 105 的集合包括服务器、工作站、台式计算机或膝上型计算机,它们的已安装软件根据扩大产品定义和相关签名的可用库的目的而被分析。

[0023] 系统管理服务器 115 控制对互连的端点 105 的可能很大的集合的管理操作,由此避免了物理上访问每个计算机的需要。此服务器和底层的系统管理基础设施被用于根据收集由分析应用所需的原始数据的目的而执行对所管理计算机的扫描操作;以及用于将扫描器在本地生成的数据移动到中央文件服务器 301,其中这些数据可以被存储在服务器的文件系统上。扫描操作可以在所有已管理端点 105 上或在特定子集上执行,后者可能由它们的角色(服务器、工作站、台式计算机或膝上型计算机)或者由运行的操作系统所确定。

[0024] 根据本发明的优选实施例,通过使用 IBM 的 Software Signature Analyzer Tool 实现上述数据收集方法;但是,本领域技术人员可以理解,产生类似结果的其它方法可以作为替代被使用。本发明目的是使用所收集的信息来产生健壮签名,其避免了现有技术的缺陷。如上所述,健壮签名不需要被更新以反映由针对相关联产品而发布的软件更新和修复所导致的改变,同时它仍旧避免了“假阳性”和“假阴性”。从所选择的计算机收获的数据存储在中央文件服务器的文件系统中,通过所述中央文件服务器的文件系统,数据可以被分析应用 303 访问。分析应用 303 实现了导致新产品定义和相关签名的创建的数据转换和分析。

[0025] 分析应用 303 支持足迹(多个文件组)的标识,所述足迹表示软件产品的已安装

实例。每个足迹需要关联于产品定义,所述产品定义通常可以从所收获的数据自动得出,或者可以由应用通过用户输入所生成。分析应用 303 将如上所述收集的数据与操作系统注册表 305 中包含的注册表信息 相结合。该已结合的产品定义和相关签名接着被生成以及存储于软件知识库 307 中。软件知识库 307 是保存整理后的产品定义和签名的集合的数据库。数据库最初可以是空的,或者它可以用由软件签名的提供者所创建的数据来填充。由分析应用产生的新定义在此数据库中进行整理。

[0026] 签名定义基于包括文件信息和注册表信息的可用信息被生成。

[0027] 在所述过程的第一阶段中,签名的基于文件部分被构建。

[0028] 文件名称以及它们相对于公共父文件夹的路径从产品的重构足迹的内容中获得,每个足迹表示:通过其可选文件已经被过滤掉的产品安装树的文件夹。产品安装树通常包括在公共根下的多个文件夹,因此它可以由一组若干足迹所表示。有时,在产品树中的整个文件夹的存在可以是可选的。这将由相对于其父文件夹的较少数量的实例所展现,因此对应于可选文件夹的足迹被清除。此过程产生了针对产品的给定版本 - 发布 (version-release) 级的“核心”文件集合的标识。通过比较对应于相同版本的足迹的内容以及相同产品的不同发布级,其名称出现在所有对应足迹中的文件可以被认为是适合于标识该版本的已安装实例,而不管发布级。理想地,将试图建立覆盖所有可能的发布级的单一签名。不过,可能出现的是:对于任意发布为公共的文件仅是所有文件的一小部分,在此情形中,可能必须创建跨发布子集的两个或多个同类文件组。进行划分的另一原因在于,万一在从一个发布到另一发布时,给定足迹在安装树中具有不同位置,从而相对路径无法被唯一地分配至其针对所有发布的文件。

[0029] 在所述过程的第二阶段中,签名的基于注册表部分被构建。

[0030] 所收获的信息包括与所安装产品相关的以及与文件信息相关的注册表数据。注册表数据可以与产品文件信息相关联的方式取决于在被收获的计算机上运行的操作系统的特性。该关联容易地在大多数 UNIX 操作系统上实现,其中平台注册表支持关于关联于已注册产品的文件的查询。在其它平台上,应用在平台注册表中注册每个实例的重要属性(比如安装位置和版本)是常见的。不幸的是,这些平台不提供查询关于关联于产品的文件的注册表的标准形式。不过,所收获的信息可包括来自其它源的数据,即,由软件安装者在安装者特定注册表中创建的注册表条目。如果安装路径是由安装者注册的信息的一部分,则注册表条目可以与对应的已安装实例相关。可替换地,如果没有适当的注册表源与所收获的数据中的已安装实例相关,则注册表信息可由用户输入。在任一情形中,在此阶段中创建的签名信息必须足以用于支持对指定软件注册表的查询,所述软件注册表由签名评估器所实现,其返回已安装实例的版本。签名的基于注册表部分还包括注册表查询的规范,其返回已安装实例的安装位置。此信息的存在允许评估器执行对签名的基于文件部分的“即时”验证,而无需对文件系统的完全扫描。

[0031] 现在考虑图 4,通过方法 400 表示了根据本发明实施例的生成产品签名的逻辑流。所述方法开始于黑色开始圈 401。继续到框 403,在全部端点 105 中收集数据。在步骤 405,所获得的数据接着被分析,并且在步骤 407 核心文件集合被标识。该核心集合表示相关的软件产品。除非产品定义已经存在(步骤 409),否则基于注册表的信息被获得(步骤 411)并与基于文件的信息(以上获得的核心集合)相结合用于产品签名的生成(步骤 413)。所

述过程在黑圈 415 结束。

[0032] 如果程序和数据以不同方式构成,如果其它模块或功能被提供,或者如果信息被存储在等价的存储器结构中,类似的考虑也可以应用。

[0033] 如果所述方法包括等价或附加步骤,类似的考虑也可以应用。

[0034] 可替换地,例如,同样取决于其上安装了系统的操作系统,获得关于所检查的端点的信息的方法可以被使用。

[0035] 尽管以上已经具体参考本发明优选实施例在某种程度上描述了本发明,应该理解,在形式和细节上的多种改变以及其它实施例是可能的。具体地,以下是明确地打算的:基本上以相同方式执行相同功能从而实现相同结果的那些单元和 / 或方法步骤的组合处于本发明的范围内。

[0036] 在任意情形中,本发明的方法可以用于发现、清点、或计量任意类型的软件产品(诸如视频游戏、多媒体工作等)。

[0037] 此外,所述程序可以被分布在任意其它计算机可读媒体(诸如一个或多个 DVD)上;可替换地,所述程序被预先加载到硬盘上、被传送到计算机、被广播、或者更广泛地被用于被直接加载到计算机的工作存储器中的任意其它形式提供。

[0038] 此外,对于本领域技术人员将很明显,提供进一步优势的附加特征并非执行本发明所必需的,并且可以被忽略或用不同特征来替换。

[0039] 在任意情形中,根据本发明的方法同样适用于通过硬件结构而执行(例如,集成到半导体材料的芯片中),或者通过软件和硬件的结合而执行。

[0040] 无疑,为了满足本地和特定需求,本领域技术人员可以将许多修改和变体应用于上述方案,然而,所有这些修改和变体被包括在由以下权利要求所限定的本发明的保护范围内。

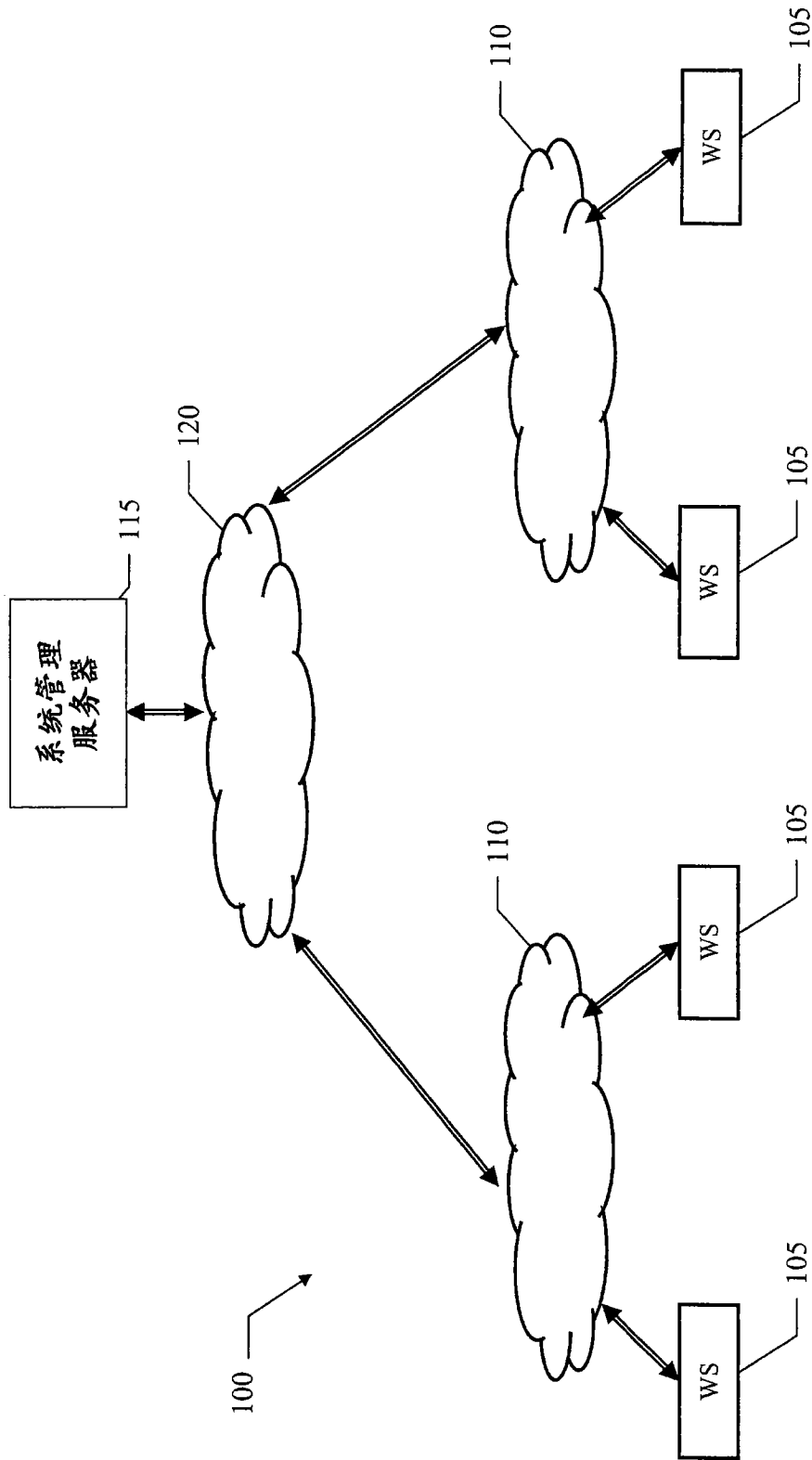


图1

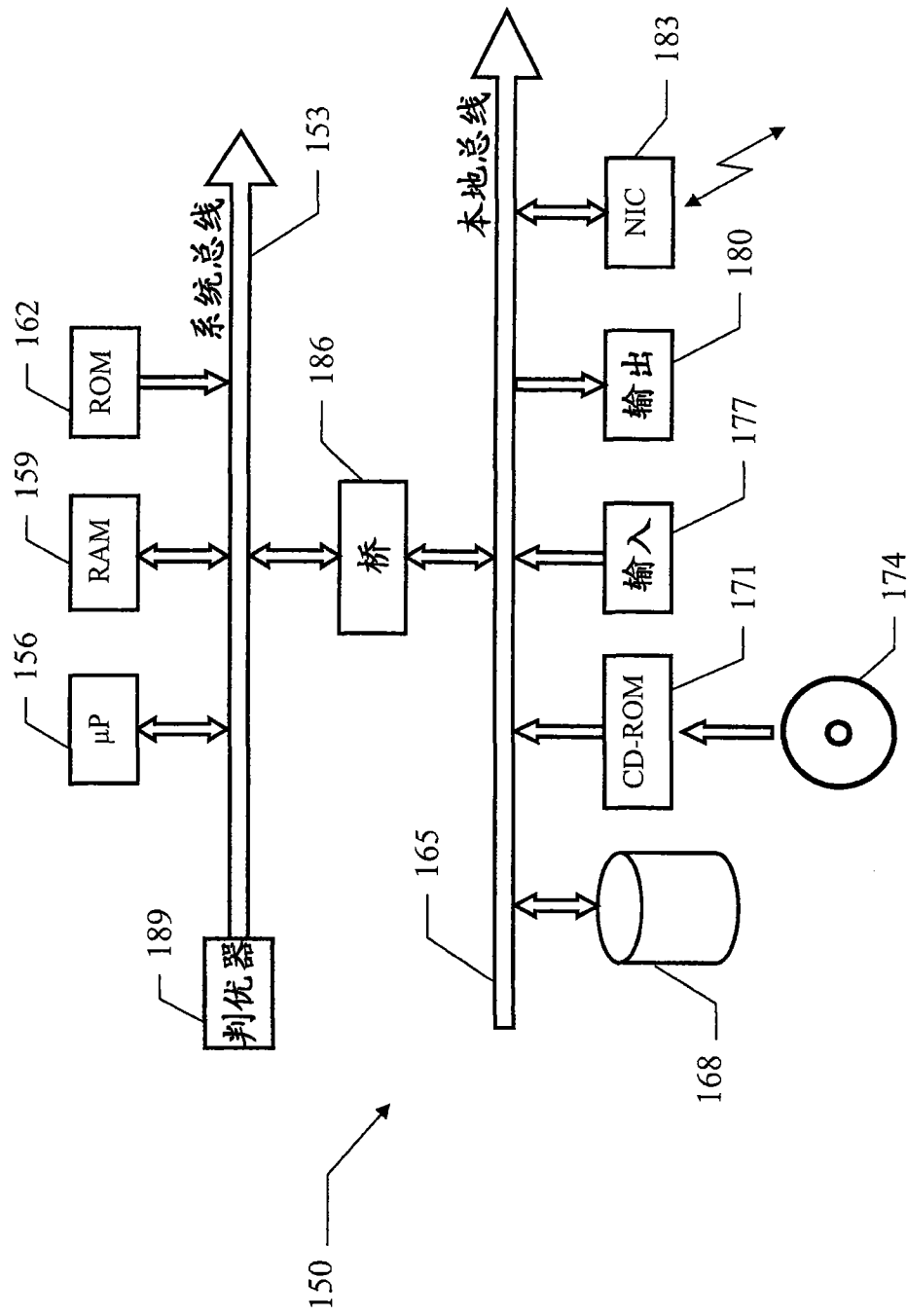


图 2

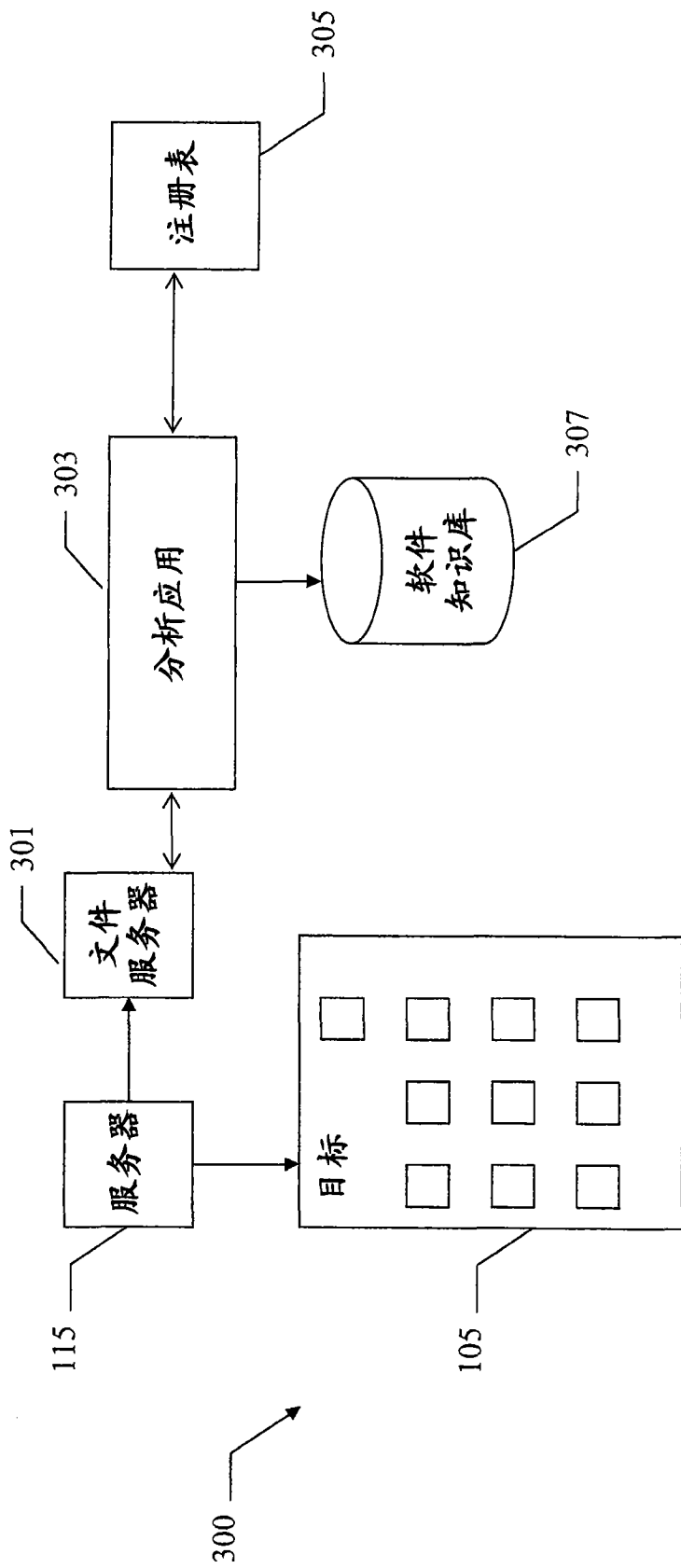


图 3

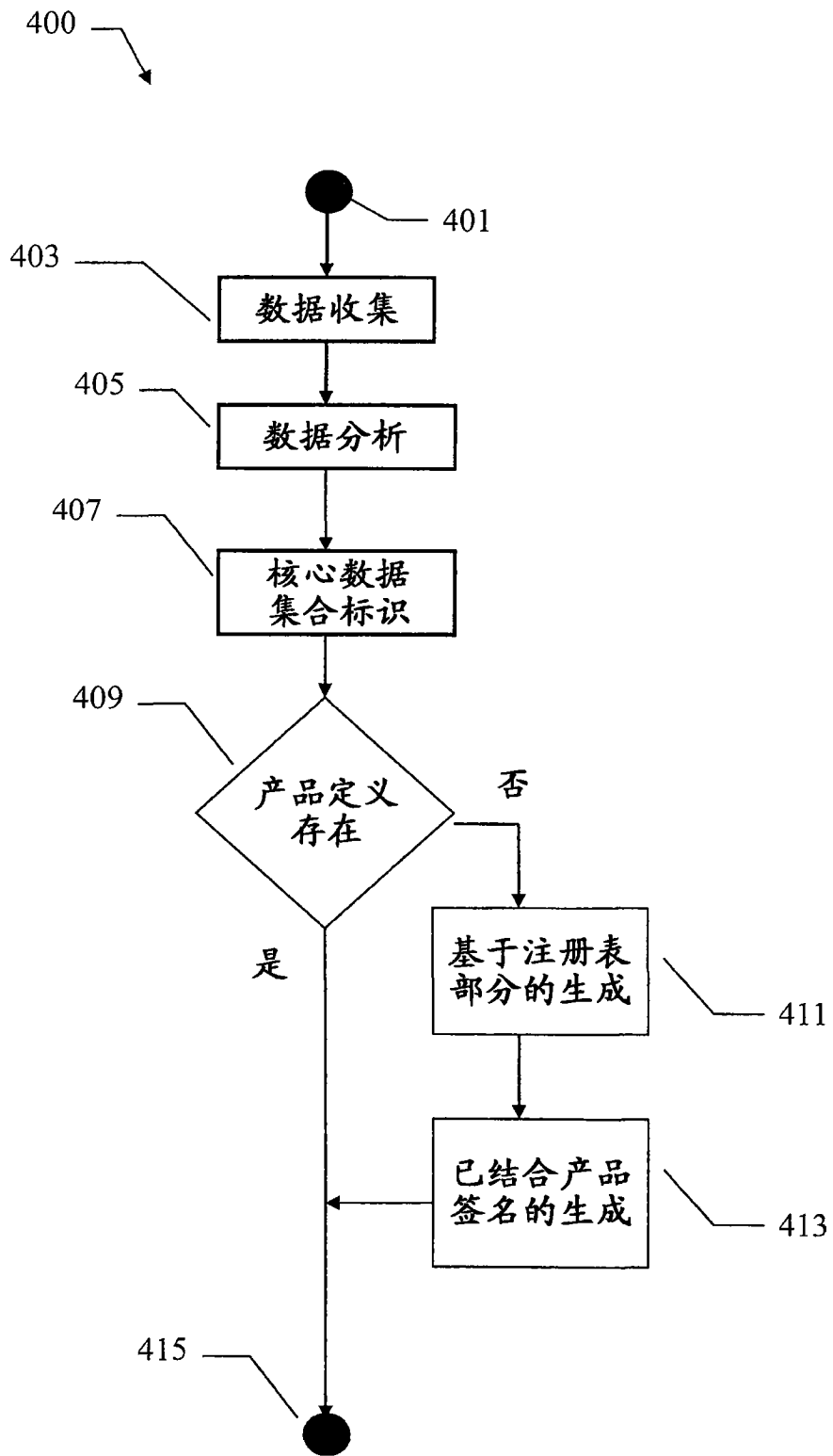


图 4