(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102067148 B
(45)授权公告日 2014.08.20

(21)申请号 200980123711.5
(22)申请日 2009.06.16
(30)优先权数据
12/144,189 2008.06.23 US
(85)PCT国际申请进入国家阶段日
2010.12.22
(86)PCT国际申请的申请数据
PCT/US2009/047505 2009.06.16
(87)PCT国际申请的公布数据
WO2009/158239 EN 2009.12.30
(73)专利权人 赛门铁克公司
地址 美国加利福尼亚州
(72)发明人 S·库利 P·维尔乔恩
(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
代理人 周亚荣 安翔
(51)Int.Cl.
*G06F 21/00*(2013.01)
(56)对比文件
US 2008/0147612 A1,2008.06.19,说明书第1-25段.
US 2002/0073328 A1,2002.06.13,全文.
US 6530024 B1,2003.03.04,全文.
CN 1567321 A,2005.01.19,全文.
审查员 康凯

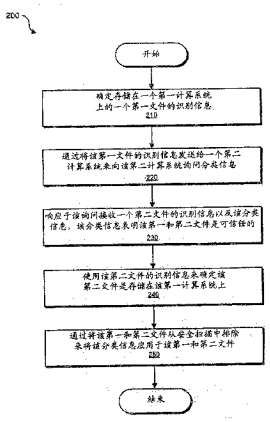
权利要求书3页 说明书13页 附图7页

(54)发明名称

用于确定文件分类的方法及系统

(57)摘要

一种用于确定文件分类的方法。该方法可以包括确定存储在一个第一计算系统上的一个第一文件的识别信息。该方法还可以包括通过将该第一文件的识别信息发送给一个第二计算系统来向该第二计算系统询问分类信息。该第一计算系统可以响应于该询问接收一个第二文件的识别信息。该第一计算系统还可以接收该分类信息。该分类信息可以表明该第一文件和第二文件是可信任的。该第一计算系统可以使用该第二文件的识别信息来确定该第二文件是存储在该第一计算系统上。该第一计算系统还可以通过将该第一和第二文件从安全扫描中排除来将该分类信息应用于该第一和第二文件。




CN 102067148 B

1. 一种用于确定文件分类的计算机实施的方法，该计算机实施的方法包括：

确定存储在一个第一计算系统上的一个第一文件的识别信息，其中该识别信息识别存储该第一文件的目录；

通过将该第一文件的识别信息发送给一个第二计算系统向该第二计算系统询问分类信息；

响应于该询问：

接收该分类信息，该分类信息指示该第一文件和第二文件是可信任的，以及

接收一个第二文件的识别信息，该第二文件被该第二计算机系统识别为与存储该第一文件的该目录相关联；

使用该第二文件的识别信息来确定该第二文件是被存储在该第一计算系统上；

通过将该第一和第二文件从安全扫描中排除来将该分类信息应用于该第一和第二文件。

2. 如权利要求 1 所述的计算机实施的方法，其中：

确定该第一文件的识别信息包括计算该第一文件的一个摘要；

该第一文件的识别信息包括该第一文件的摘要；

该第一文件的识别信息包括存储该第一文件的该目录的一个名称。

3. 如权利要求 2 所述的计算机实施的方法，其中：

该摘要包括该第一文件的一个散列；

存储该第一文件的目录的名称包括一个标准化的目录路径。

4. 如权利要求 1 所述的计算机实施的方法，其中：

该第二文件的识别信息包括该第二文件的一个文件名以及该第二文件的一个摘要；

使用该第二文件的识别信息来确定该第二文件是存储在该第一计算系统上包括：

确定该第一计算装置包括具有该第二文件的文件名的一个文件；

确定该第二文件的摘要与具有该文件名的文件的摘要相匹配。

5. 如权利要求 4 所述的计算机实施的方法，其中：

该第二文件的识别信息包括该第二文件的大小；

使用该第二文件的识别信息来确定该第二文件是存储在该第一计算装置上包括：

确定具有该第二文件的文件名的文件与该第二文件是大小相同的。

6. 如权利要求 1 所述的计算机实施的方法，其中：

该第一文件已被一个第一软件程序安装在该第一计算装置上；

该第二文件已被该第一软件程序安装在该第一计算装置上。

7. 如权利要求 1 所述的计算机实施的方法，进一步包括：

接收对于一组文件中的每个文件的识别信息，其中：

该组文件是与安装了该第一文件的一个应用程序相关联的；

该组文件包括该第二文件；

接收对于该组文件中的每个文件的分类信息。

8. 如权利要求 7 所述的计算机实施的方法，进一步包括：

确定对于一个文件子组中的每个文件的识别信息，其中：

该文件子组包括该第一文件；

该文件子组是与该应用程序相关联的该组文件的一个子组；

通过将该文件子组中的每个文件的识别信息发送给该第二计算系统来向该第二计算系统询问分类信息。

9. 一种用于确定文件分类的计算机实施的方法，该计算机实施的方法包括：

从一个第一计算系统接收存储在该第一计算系统上的一个第一文件的识别信息，其中该识别信息识别存储该第一文件的目录；

使用该识别信息在一个分类数据库中搜索对于该第一文件的分类信息，该分类数据库是存储在一个第二计算系统上；

识别与存储该第一文件的该目录相关联的一个第二文件；

将该第二文件的识别信息发送到该第一计算系统上；

将对于该第一文件的分类信息以及对于该第二文件的分类信息发送到该第一计算系统上，其中所述第一计算系统使用所述第二文件的所述识别信息来确定所述第二文件是被存储在所述第一计算系统上，并且通过从安全扫描中排除所述第一文件和所述第二文件来将所述分类信息应用于所述第一文件和所述第二文件。

10. 如权利要求 9 所述的计算机实施的方法，其中：

该第一文件的识别信息包括该第一文件的一个摘要；

该第一文件的识别信息包括存储该第一文件的该目录的一个名称。

11. 如权利要求 10 所述的计算机实施的方法，其中：

该摘要包括该第一文件的一个散列；

存储该第一文件的目录的名称包括一个标准化的目录路径。

12. 如权利要求 9 所述的计算机实施的方法，其中：

该第一文件已被一个第一软件程序安装在该第一计算装置上；

使用存储该第一文件的目录的名称来识别该第二文件以便发现与该第一软件程序相关联的一个目录；

识别该第二文件包括确定与该第一软件程序相关联的目录包括该第二文件。

13. 如权利要求 9 所述的计算机实施的方法，其中：

该第二文件的识别信息包括以下至少一项；

该第二文件的一个文件名；

该第二文件的大小；

该第二文件的一个摘要。

14. 如权利要求 9 所述的计算机实施的方法，进一步包括：

发送对于一组文件中的每个文件的识别信息，其中：

该组文件是与安装了该第一文件的一个应用程序相关联的；

该组文件包括该第二文件；

接收对于该组文件中的每个文件的分类信息。

15. 一种用于确定文件分类的系统，所述系统包括：

识别模块，所述识别模块用于：

确定存储在一个第一计算系统上的一个第一文件的识别信息，其中该识别信息识别存储该第一文件的目录；

搜索模块，所述搜索模块用于：

通过将该第一文件的识别信息发送给一个第二计算系统来向该第二计算系统询问分类信息；

响应于该询问：

接收该分类信息，该分类信息指示该第一文件和第二文件是可信任的，以及

接收一个第二文件的识别信息，该第二文件被该第二计算机系统识别为与存储该第一文件的该目录相关联；

应用模块，所述应用模块用于：

使用该第二文件的识别信息来确定该第二文件是存储在该第一计算系统上；

通过将该第一和第二文件从安全扫描中排除来将该分类信息应用于该第一和第二文件。

# 用于确定文件分类的方法及系统

## 背景技术

[0001] 消费者们和企业越来越多地依赖计算机来存储敏感数据。因此，恶意的程序员们似乎不断地增加他们的努力以便获取对其他人的计算机的非法控制和访问。带有恶意动机的计算机程序员们已经创建了并且继续创建着病毒、特洛伊木马程序、蠕虫、以及意欲危害属于他人的计算机系统和数据的其他程序。这些恶意程序通常被称为恶意软件。

[0002] 安全软件公司正在通过定期地为他们的客户创建并分发恶意软件签名（例如，识别恶意软件的散列）来与增长的恶意软件势头做斗争。例如，安全软件公司可以一天多次向他们的客户发送恶意软件签名更新。通过经常地对恶意软件签名进行更新，安全软件公司可以帮助他们的客户保护他们的计算机免受新的并且变化的威胁。

[0003] 每当一个客户接收到一个恶意软件定义更新时，该客户的计算机可能需要对大量文件进行重新扫描以便确保没有恶意软件在该计算机上运行。因此，客户们的计算机在每次收到一个恶意软件签名更新时都可能受到一次性能减损 (performance hit)。在客户计算机以及网络上的性能损失随着签名更新的大小和频率的增加而增加。这种性能损失可能导致一种负面的客户体验。

[0004] 安全软件开发人员已经尝试通过跳过已知的良好文件（例如，已知没有恶意软件的文件）来减少进行安全扫描所要求的时间和网络流量。在跳过一个文件之前，安全软件典型地必须识别该文件是合法的并且没有恶意软件。安全软件开发人员已经实施了至少两种不同的方法来识别已知的良好文件以便减少安全扫描的次数。

[0005] 在第一种方法中，一个客户端机器可以保持多个已知良好文件的散列的一个数据库。当进行一次恶意软件扫描时，该客户端机器可以询问该数据库以便识别已知的良好文件。然后，该客户端机器可以跳过这些已知文件，这可以允许更快地完成扫描。然而，保持已知良好文件的散列的一个数据库可能并不理想。维持这种数据库可能要求经常的更新，而这些更新会增加网络流量。同样，该数据库可能变大并且也许并不提供所希望的效率。

[0006] 在第二种方法中，一个客户端机器可以对一个驱动器进行扫描。该客户端机器可以计算存储在这个驱动器上的多个文件的散列并且可以将这些散列发送到一个服务器上。然后，该服务器可以确定这些散列是否对应于已知的良好文件。这个技术同样有缺点。将多个文件散列发送给一个服务器可能产生不必要的客户端 - 服务器通信量并且可能耗费太多的网络带宽。另外，许多用户互联网连接是非对称的，其中上传带宽比下载带宽要小得多。因此，将来自用户的计算机的多个散列上传到一个安全软件服务器可能是一个缓慢的、耗费资源的过程。

[0007] 相关申请的交叉引用

[0008] 本披露涉及于 2008 年 5 月 30 日提交的题为“针对恶意软件对文件进行扫描的方法及系统”的美国申请号 12/130,559 以及于 2008 年 5 月 30 日提交的题为“用于确定一个文件组的系统及方法”的美国申请号 12/130,839，这两个申请各自的披露通过引用以其全文结合在此。

## 发明内容

[0009] 本披露的多个实施方案是指向通过跳过可信任的文件（即，已知的良好文件）来增加安全扫描速度。一个客户端装置（例如，一个第一计算系统）上的安全模块可试图在该客户端装置上识别多个可信任的文件。在识别多个可信任文件的过程中，该安全模块可以确定存储在该客户端装置上的一个第一文件的识别信息。该安全模块可以通过计算该第一文件的一个摘要来确定识别信息。该安全模块还可以通过识别存储该第一文件的一个目录的名称来确定识别信息。该安全模块可以通过将该第一文件的识别信息发送给一个服务器（例如，一个第二计算系统）来向该服务器询问分类信息。

[0010] 该服务器可以使用该识别信息在一个分类数据库中搜索对于该第一文件的分类信息。例如，该服务器可以在该分类数据库中搜索与该第一文件的摘要相匹配的一个文件摘要。除了搜索文件摘要之外或者替代搜索文件摘要，该服务器可以在该分类数据库中搜索与包括在识别信息中的目录名相匹配的一个目录名。该服务器还可以发现与该目录名相关联的一个或多个额外文件。然后，该服务器可以将这些额外文件的识别信息发送给客户端装置。该服务器还可以将对于该第一文件以及这些额外文件的分类信息发送给该第一计算装置。

[0011] 该分类信息可以表明该第一文件以及这些额外文件是否是可信任的。如果这些文件是可信任的，那么在该第一计算装置上的安全软件可以将这些文件从安全扫描中排除。通过排除可信任的文件，安全扫描可以更有效率地运行。另外，本披露的多个实施方案可以减少需要用来识别可信任文件的网络带宽量（特别是客户端侧的上传带宽），因为一个客户端装置可以仅需要上传一个文件摘要来获取对于多个文件的分类信息。在其他的实施方案中，客户端装置可以将几个文件摘要发送给服务器，并且服务器可以返回对于那些文件以及额外文件的分类信息。

[0012] 来自上述任一个实施方案的多种特征可根据在此说明的通用原理彼此相互结合使用。通过阅读以下的详细说明连同附图和权利要求，将会更加全面地理解这些以及其他的实施方案、特征、以及优点。

## 附图说明

[0013] 这些附图展示了多个示例性实施方案并且是本说明书的一部分。这些附图与以下的说明一起展现并解释了本披露的不同原理。

[0014] 图 1 是用于实施本披露的多个实施方案的一种示例性网络的框图。

[0015] 图 2 是根据某些实施方案用于确定文件分类的一个示例性方法的流程图。

[0016] 图 3 是根据某些实施方案用于确定文件分类的另一个示例性方法的框图。

[0017] 图 4 是根据某些实施方案的一个示例性分类信息数据库的框图。

[0018] 图 5 是根据某些实施方案在客户端与服务器之间的示例性通信的流程图。

[0019] 图 6 是一个示例性计算系统的框图，该系统能够实施在此说明和 / 或展示的这些实施方案中的一个或多个。

[0020] 图 7 是一个示例性计算网络的框图，该网络能够实施在此说明和 / 或展示的这些实施方案中的一个或多个。

[0021] 贯穿这些附图，相同的参考字符以及说明表示相似的但并不一定完全相同的要

素。虽然在此说明的这些示例性实施方案可容许进行不同的修改以及多种替代形式，在此仍在附图中以举例的方式示出多个具体的实施方案并且对其进行了详细的说明。然而，在此说明的多个示例性实施方案并非旨在被限于所披露的这些具体形式。相反，本披露覆盖落入所附权利要求范围内的所有修改、等效物、以及替代方案。

## 具体实施方式

[0022] 如以下将更详细说明的，本披露总体上涉及通过将对于一个或多个文件的识别信息发送给一个服务器来确定对于多个文件的分类信息的系统及方法。如以下详细说明的，客户端装置上的安全软件可以对于一个第一文件的分类信息询问服务器上的一个分类数据库。响应于该请求，服务器可以向客户端装置发送对于该第一文件的分类信息。服务器还可以向客户端装置发送对于与该第一文件相关联的一个或多个额外文件的识别信息和分类信息。客户端装置可以使用该分类信息来确定该客户端装置上的文件是否是可信任的以及是否可以在安全扫描中跳过。本披露的多个实施方案可以减少客户端的上游带宽并且使在客户端系统上对文件进行分类所需要的网络往返请求的数目最小化。本披露的多个实施方案还提供了不同的其他特征和优点。

[0023] 图 1 示出了具有客户端 120 和服务器 150 的一个网络 100。客户端 120 和服务器 150 可以在网络 140 上进行通信。网络 140 可以是互联网、局域网 (LAN)、广域网 (WAN)、或任何其他适当的计算机网络。客户端 120 可以是台式计算机、膝上计算机、移动计算装置、网络计算装置、或任何其他适当的计算装置。在一些实施方案中，客户端 120 可以被称为一个第一计算装置。客户端 120 可以包括安全软件 122、可信任文件 124、应用程序 126、以及目录 128。

[0024] 安全软件 122 可以是被编程来确定一个文件的识别信息并且向一个第二计算系统询问对于该文件的分类信息的任何模块、应用程序、或其他计算机可执行代码。安全软件 122 可以确定客户端 120 上的哪些文件是可信任的，这样使得安全软件 122（或任何其他应用程序）可以在安全扫描中跳过这些信任的文件。为了确定客户端 120 上的文件是否是可信任的，安全软件 122 可以识别目录 128 中的一个第一文件，如文件 130。目录 128 可以是客户端 120 上的一个文件系统中的一个目录。

[0025] 在识别了文件 130 之后，安全软件 122 可以确定对于文件 130 的识别信息。例如，安全软件 122 可以计算对于文件 130 的一个摘要。在其他的实施方案中，安全软件 122 可以识别出对于文件 130 的一个预先计算的摘要。安全软件 122 还可以确定目录 128 的一个目录名。在一些实施方案中，目录 128 的目录名可以是对于目录 128 的一个标准化的目录路径。通过确定对于文件 130 的识别信息，安全软件 122 也许能够通过将识别信息而不是整个文件 130 发送给服务器 150 来请求对于文件 130 的分类信息。典型地，对于一个文件的识别信息可以比文件本身小。因此，发送对于一个文件的识别信息比发送文件本身耗费更少的网络带宽。

[0026] 安全软件 122 可以将对于文件 130 的识别信息发送给服务器 150 以便向分类信息数据库 152 询问对于文件 130 的分类信息。图 4 以及相应的说明提供了分类数据库以及可以如何创建分类数据库的一个实例。服务器 150 可以在分类信息数据库 152 中搜索与文件 130 相关联的分类信息。在一些实施方案中，服务器 150 在分类信息数据库 152 中搜索与

文件 130 的摘要相匹配的一个摘要。服务器 150 可以发现摘要 162 与文件 130 的摘要相匹配。

[0027] 在一些实施方案中，在分类信息数据库 152 中搜索与文件 130 的摘要相匹配的一个摘要之前，服务器 150 可以在分类信息数据库中搜索与目录 128 的目录名相匹配的一个目录名。例如，服务器 150 可以确定目录名 160 与目录 128 的目录名相匹配。在分类信息数据库 152 中识别了与目录 128 的目录名相匹配的一个目录之后，服务器 150 可以搜索与一个摘要的目录相关联的文件，该摘要与文件 130 的摘要相匹配。

[0028] 分类信息数据库 152 可以识别与文件 130 相关联的多个额外的文件摘要。例如，分类信息数据库 152 可以识别文件摘要 164 和 166，因为文件摘要 164 和 166 同样是与目录名 160 相关联的。分类信息数据库 152 可以将摘要 164 和 166 发送给客户端 120 上的安全软件 122。分类信息数据库 152 还可以将分类信息 170、172、和 174 发送给安全软件 122。

[0029] 在接收了对于这些额外文件的识别信息之后，安全软件 122 可以确定这些额外文件是否被包括在目录 128 中。例如，安全软件 122 可以确定摘要 164 和 166 是否与文件 132 和 134 的摘要相匹配。作为这个确定过程的一部分，安全软件 122 可以计算对于文件 132 和 134 的摘要。在其他的实施方案中，安全软件 122 可以识别对于文件 132 和 134 的预先计算的摘要。

[0030] 安全软件 122 可以确定摘要 164 对应于文件 132 并且摘要 166 对应于文件 134。然后，安全软件 122 可以将分类信息 172 应用于文件 132 并且将分类信息 174 应用于文件 134。安全软件 122 还可以将分类信息 170 应用于文件 130。如果分类信息 170、172、和 174 表明文件 130、132、和 134 是可信任的，那么安全软件 122 可以通过将这些文件从安全扫描中排除将分类信息 170、172、和 174 应用于文件 130、132、和 134。尽管图 1 示出了对于与一个目录相关联（例如，存储在其中）的每个文件的分类信息，在一些实施方案中在一个目录中的每个文件可以是与相同的分类信息相关联的。

[0031] 图 1 示出了客户端 120 可以包括一个应用程序 126。目录 128 可以是一个文件夹，该文件夹是由应用程序 126 安装和 / 或以其他方式与应用程序 126 相关联。分类信息数据库 152 可以包括对于由应用程序 126 安装的文件的分类信息。例如，安全模块可以监测应用程序 126 的安装以便确定哪些文件是由应用程序 126 安装的。然后，安全模块可以对由应用程序 126 安装的文件进行分析和分类。对应于图 4 的披露讨论了对多个文件进行分析和分类的过程。

[0032] 图 2 是一个流程图，展示了可以由一个客户端装置（如客户端 120）进行的多个步骤。客户端装置上的安全软件可以确定存储在一个第一计算系统上的一个第一文件的识别信息（步骤 210）。确定一个第一文件的识别信息可以包括计算对于该第一文件的一个摘要。摘要可以是一个文件识别符，如一个散列。一个散列可以是一个值、代码、和、或使用一个散列函数（例如，将任意长度的数据流作为输入而产生某个固定大小的值的输出的一个函数）计算的其他数据。这种散列函数可以是循环冗余校验 (CRC) 散列函数。在其他的实施方案中，散列函数可以是加密散列函数，如消息摘要算法 5(MD5) 散列函数。这种散列函数还可以是安全散列算法 (SHA) 加密散列函数，如 SHA-0、SHA-1、SHA-256、或 SHA-512 散列函数。还可以通过将任何其他公开的或专有的散列算法应用于该第一文件来计算该第一文件的散列。

[0033] 确定该第一文件的识别信息还可以包括确定存储该第一文件的一个目录。例如，安全软件可以确定存储该第一文件的目录的目录名。该目录名可以是该目录的一个文件路径。在一些实施方案中，目录名可以是一个标准化的目录路径，这种标准化的目录路径是通过将用于一种特定机器的绝对目录路径转换为消除了该目录路径的机器专有部分的通用目录路径来创建。换言之，可以通过移除专用于一个客户端机器的一个目录路径中的数据但保留该目录路径中对于可能安装该目录的任何机器都是通用的部分来创建一个标准化的目录路径。

[0034] 在确定了该第一文件的识别信息之后，安全软件可以向一个第二计算装置（例如，服务器）询问分类信息（步骤 220）。安全软件可以通过将该第一文件的识别信息发送给该第二计算系统来询问该第二计算系统。例如，安全软件可以将对于该第一文件的摘要和目录名发送给该第二计算系统上的一个分类信息数据库。

[0035] 该第一计算系统可以响应于该询问接收一个第二文件的识别信息（步骤 230）。该第二文件的识别信息可以是该第二文件的一个文件名、该第二文件的大小、该第二文件的一个摘要、和 / 或对于该第二文件的任何其他识别信息。该第一计算系统还可以接收对于该第一和第二文件的分类信息（步骤 230）。分类信息可以是指定该第一和第二文件是否是可信任的分类信息。在一些实施方案中，分类信息可以是与这两个文件相关联的表明这些文件是否可信任的一个单一的分类。在其他的实施方案中，该第一计算装置可以接收对于该第一和第二文件各自不同的分类信息。

[0036] 安全软件可以使用第二文件的识别信息来确定该第二文件是存储在该第一计算系统上（步骤 240）。使用该第二文件的识别信息来确定该第二文件是存储在该第一计算装置上可以包括确定该第一计算装置包括具有该第二文件的文件名的一个文件以及确定该第二文件的摘要与具有该第二文件的文件名的文件摘要相匹配。换言之，该安全软件可以首先检查第一计算装置以便确定该第一计算装置是否具有一个文件，该文件具有的名称与从第二计算系统接收的名称相匹配。如果安全软件发现了一个文件名的匹配，则安全软件可以将该第一计算装置上的文件的散列与从该第二计算装置接收的散列进行比较。如果这些散列相匹配，则安全软件可以确定该第一计算装置上的文件是由该第二计算装置识别的文件（即，该第一计算装置包括该第二文件）。

[0037] 在一些实施方案中，该第二文件的识别信息包括该第二文件的大小。在此类实施方案，使用该第二文件的识别信息来确定该第二文件被存储在该第一计算装置上可以包括确定具有该第二文件的文件名的文件与该第二文件是大小相同的。换言之，安全软件可以通过确定该文件与该第二文件是大小相同的、通过确定这些文件的散列函数相匹配、和 / 或通过确定这些文件具有相同的名称来确定该文件就是该第二文件。

[0038] 安全软件可以通过将该第一和第二文件从安全扫描中排除来将分类信息应用于该第一和第二文件（步骤 250）。例如，安全软件可以包括在一个可信任文件的列表中的第一和第二文件。当进行安全扫描时，安全软件可以跳过被识别为在该可信任文件列表中的文件。

[0039] 在一些实施方案中，步骤 210（确定一个第一文件的识别信息）可以包括确定多个文件的识别信息。例如，安全软件可以确定一个目录中的几个文件（或很多文件）的识别信息，其中该目录包括一组文件。因此，安全软件可以确定对于该目录中的一个文件子组的

识别信息，并且该文件子组可以包括该第一文件。在此类实施方案中，步骤 220 可以包括将该文件子组中的每个文件的识别信息发送给该第二计算系统。该第二计算系统也许能够通过从该组文件接收多个文件而不是仅从该组文件接收一个单一的文件而更快速并且更准确地确定一组相关联的文件中的其他文件。

[0040] 在一些实施方案中，步骤 230（接收一个第二文件的识别信息）可以包括接收对于一组文件的识别信息。该组文件可以包括该第二文件。该组文件可以包括与存储该第一文件的目录相关联的每个文件。

[0041] 图 3 示出了用于确定文件分类的一种计算机实施的方法的流程图。图 3 中示出的这些步骤可以由与正在请求对于多个文件的分类信息的一个客户端进行通信的服务器（例如，一个第二计算系统）来执行。该第二计算系统可以从一个第一计算系统接收一个第一文件的识别信息（步骤 310）。该第一文件可以是存储在该第一计算系统上的一个文件。该第二计算系统可以使用该识别信息在一个分类数据库中搜索对于该第一文件的分类信息（步骤 320）。该分类数据库可以存储在该第二计算系统上。如先前所讨论的，该第二计算系统可以通过在该分类数据库中搜索与该第一文件的摘要和 / 或目录相匹配的一个摘要和 / 或目录而在该分类数据库中搜索该第一文件的分类信息。

[0042] 在该分类数据库中对该第一文件进行识别之后，该第二计算系统可以识别与该第一文件相关联的一个第二文件（步骤 330）。该第二计算系统可以通过搜索与该第一文件的相同的目录相关联的多个文件来识别与该第一文件相关联的多个文件。换言之，如果该第一文件与第二文件存储在同一个目录中，那么该第一文件可能是与该第二文件相关联的。

[0043] 该第二计算系统可以将该第二文件的识别信息发送给该第一计算装置（步骤 340）。如先前所提及的，该第二文件的识别信息可以包括该第二文件的一个摘要、该第二文件的一个名称、该第二文件的大小、和 / 或识别该第二文件或可以使该第二文件与其他文件相区别的任何其他信息。该第二计算系统可以将对于该第一文件的分类信息以及对于该第二文件的分类信息发送给该第一计算装置（步骤 350）。

[0044] 尽管步骤 340 和 350 被示出为分离的步骤，该第二计算装置可以同时进行步骤 340 和 350。换言之，该第二计算装置可以同时发送识别信息和分类信息。在一些实施方案中，该第二计算装置可以识别与该第一文件相关联的一组文件，并且该组文件可以包括该第二文件。该第二计算装置可以将对于该组文件中的每个文件的识别信息和分类信息发送给该第一计算装置。

[0045] 图 4 示出了一个分类信息数据库 400。分类信息数据库 400 可以将多个目录与多个文件相关联。分类信息数据库 400 可以通过将多个目录的目录名与对于多个文件的文件摘要相关联而使得多个目录与多个文件相关联。例如，目录名 410 可以是与文件摘要 420、422、424、426、以及 428 相关联的。目录名 412 可以是与文件摘要 430、432、以及 434 相关联的。目录名 414 可以是与文件摘要 440、442、444、以及 446 相关联的。

[0046] 分类信息数据库 400 还可以将多个文件与分类信息相关联。例如，文件摘要 420 可以是与分类信息 450 相关联的、文件摘要 422 可以是与分类信息 452 相关联的、文件摘要 424 可以是与分类信息 454 相关联的、文件摘要 426 可以是与分类信息 456 相关联的、并且文件摘要 428 可以是与分类信息 458 相关联的。文件摘要 430、432、和 434 可以是与分类信息 460 相关联的。在一些实施方案中，分类信息可以是与一个目录相关联的。例如，目录

414 可以是与分类信息 470 相关联的。因此，与目录 414 相关联的任何文件还可以是与分类信息 470 相关联的。因此，分类信息 470 可以应用于文件摘要 440、文件摘要 442、文件摘要 444、以及文件摘要 446。

[0047] 当一个客户端向分类信息数据库 400 询问对于一个文件的分类信息时，分类信息数据库 400 可以搜索与该客户端上的文件相匹配的一个文件摘要。一旦分类信息数据库 400 发现了一个匹配的文件摘要，分类信息数据库 400 可以将与该摘要相关联的分类信息返回给该客户端。分类信息数据库 400 还可以返回与该客户端上的文件相关联的其他文件摘要。

[0048] 作为一个实例，一个客户端可以利用一个第一摘要来询问分类信息数据库 400。分类信息数据库 400 可以确定该第一摘要与文件摘要 424 相匹配。分类信息数据库 400 可以返回对于文件摘要 424 的分类信息 454。分类信息数据库 400 还可以向客户端返回文件摘要 420、422、426、和 428 以及它们相应的分类信息。

[0049] 可以使用一个分类工具来创建或填充分类信息数据库 152。该分类工具可以监测多个软件应用程序的安装并且对由这些软件应用程序安装的文件进行分类。例如，在识别了由一个应用程序安装的一组文件后，可以针对恶意软件对这些文件进行扫描。可以基于这些文件包含恶意软件的可能性对它们进行分类。这些文件和它们相应的分类信息可以被加载到一个面向客户端的数据库系统（如分类数据库 400）中。

[0050] 图 5 展示了一个流程图，它示出了在客户端 502 和服务器 504 之间的通信。客户端 502 可以选择一个文件（步骤 510）。然后，客户端 502 可以计算对于该文件的一个摘要（步骤 515）。客户端 502 可以将对于该文件的一个摘要和目录信息发送给服务器 504（步骤 520）。服务器 504 可以发现与该文件相关联的一个目录（步骤 525）。服务器 504 还可以识别与该目录相关联的多个额外文件（步骤 530）。服务器 504 可以将对于这些额外文件的识别信息发送给客户端 502（步骤 535）。同时（或在不同的时间），服务器 504 可以将对于原始文件以及这些额外文件的分类信息发送给客户端 502（步骤 540）。客户端 502 可以从服务器 504 接收该信息并且使用该分类信息识别多个文件（步骤 545）。然后，客户端 502 可以将该分类信息应用于这些文件（步骤 550）。

[0051] 图 6 是一个示例性计算系统 610 的框图，该系统能够实施在此说明和 / 或展示的一个或多个实施方案。计算系统 610 广义上代表能够执行计算机可读指令的任何单处理器或多处理器的计算装置或系统。计算系统 610 的多个实例包括但不限于工作站、膝上计算机、客户侧终端、服务器、分布式计算系统、手持装置、或任何其他计算系统或装置。在其最基本的配置中，计算系统 610 可以包括至少一个处理器 614 以及一个系统内存 616。

[0052] 处理器 614 总体上代表能够处理数据或解释并执行多个指令的任何类型或形式的处理单元。在某些实施方案中，处理器 614 可以从一个软件应用程序或模块中接收指令。这些指令可以致使处理器 614 执行在此所说明和 / 或展示的一个或多个示例性实施方案的功能。例如，处理器 614 可以单独地或与其他元件相结合执行和 / 或作为一种手段来执行在此说明的确定、计算、识别、应用、使用、询问、接收、和 / 或者发送步骤中的一个或多个。处理器 614 还可以执行和 / 或作为一种手段来执行在此说明和 / 或展示的任何其他步骤、方法、或过程。

[0053] 系统内存 616 总体上代表能够存储数据和 / 或其他计算机可读指令的任何类型或

形式的易失性或非易失性存储装置或媒质。系统内存 616 的多个实例包括 :但不限于,随机存取存储器 (RAM)、只读存储器 (ROM)、闪存、或任何其他适当的存储装置。尽管没有要求,在某些实施方案中计算系统 610 可以既包括一个易失性内存单元 ( 例如,系统内存 616) 又包括一个非易失性存储装置 ( 例如,如以下详细说明的主存储装置 632)。

[0054] 在某些实施方案中,示例性计算系统 610 除了处理器 614 和系统内存 616 外还可以包括一个或多个部件或元件。例如,如图 6 所示,计算系统 610 可以包括一个内存控制器 618、一个输入 / 输出 (I/O) 控制器 620、以及一个通信接口 622,它们各自均可通过一个通信基础结构 612 而互联。通信基础结构 612 总体上代表能够协助在一种计算装置的一个或多个部件之间进行通信的任何类型或形式的基础结构。通信基础结构 612 的实例包括但不限于,一条通信总线 ( 例如 ISA、PCI、PCIe、或类似总线 ) 和一个网络。

[0055] 内存控制器 618 总体上代表能够处理内存或数据或能够控制计算系统 610 的一个或多个部件之间通信的任何类型或形式的装置。例如,在一些实施方案中,内存控制器 618 可以通过通信基础结构 612 控制在处理器 614、系统内存 616、以及 I/O 控制器 620 之间的通信。在某些实施方案中,内存控制器可以单独地或与其他元件相结合来执行和 / 或作为一种手段来执行在此说明和 / 或者展示的步骤或特征中的一个或多个,如确定、计算、识别、应用、使用、询问、接收、和 / 或发送。

[0056] I/O 控制器 620 总体上代表能够协调和 / 或控制一种计算装置的输入和输出功能的任何类型或形式的模块。例如,在一些实施方案中 I/O 控制器 620 可以控制或协助在计算系统 610 的一个或多个元件 ( 如处理器 614、系统内存 616、通信接口 622、显示适配器 626、输入接口 630、以及存储接口 634) 之间的数据传送。例如,I/O 控制器 620 可以被用来单独地或与其他元件相结合执行和 / 或作为一种手段来执行在此说明的确定、计算、识别、应用、使用、询问、接收、和 / 或发送步骤中的一项或多项。I/O 控制器 620 还可用于执行和 / 或作为一种手段用于执行本披露中提出的其他步骤和特征。

[0057] 通信接口 622 广义地代表能够协助示例性计算系统 610 与一个或多个另外的装置之间进行通信的任何类型或形式的通信装置或适配器。例如,在某些实施方案中,通信接口 622 可协助计算系统 610 与包括多个另外的计算系统的一个私人或公共网络之间的通信。通信接口 622 的实例包括而不限于 :一种有线网络接口 ( 例如一个网络接口卡 )、一种无线网络接口 ( 例如一种无线网络接口卡 )、一种调制解调器、以及任何其他适当的接口。在至少一个实施方案中,通信接口 622 可通过到一个网络 ( 如互联网 ) 的一种直接链接来提供到一台远程服务器的直接连接。通信接口 622 还可以间接地提供这种连接,例如通过一个局域网 ( 如一个以太网 )、一个个人局域网、一个电话或缆线网、一种蜂窝电话连接、一种卫星数据连接、或任何其他适当的连接。

[0058] 在某些实施方案中,通信接口 622 还可以代表一种主机适配器,它被配置为用于通过一条外部总线或通信信道协助计算系统 610 与一个或多个附加网络或存储装置之间的通信。主机适配器的实例包括而不限于,SCSI 主机适配器、USB 主机适配器、IEEE 694 主机适配器、SATA 和 eSATA 主机适配器、ATA 和 PATA 主机适配器、光纤通道接口适配器、以太网适配器、或类似适配器。通信接口 622 还可以允许计算系统 610 参与分布式计算或远程计算。例如通信接口 622 可以从一个远程装置接收指令或向一个远程装置发送指令用于执行。在某些实施方案中,处理器 622 可以单独地或与其他元件相结合来执行和 / 或作为一

种手段用来执行在此披露的确定、计算、识别、应用、使用、询问、接收、和 / 或者发送步骤中的一个或多个。通信接口 622 还可以用于执行和 / 或作为一种手段用于执行本披露中提出的其他步骤和特征。

[0059] 如图 6 所示，计算系统 610 还可以包括通过一种显示适配器 626 联接到通信基础结构 612 上的至少一个显示装置 624。显示装置 624 总体上代表能够视觉上显示由显示适配器 626 发来的信息的任何类型或形式的装置。类似地，显示适配器 626 总体上代表被配置为用于传送来自通信基础结构 612（或来自一个帧缓冲器，如本领域中已知的）的图形、文本、以及其他数据用于在显示装置 624 上进行显示的任何类型或形式的装置。

[0060] 如图 6 所示，示例性的计算系统 610 还可以包括通过一个输入接口 630 联接到通信基础结构 612 上的至少一个输入装置 628。输入装置 628 总体上代表能够将计算机或者人产生的输入提供到示例性计算系统 610 上的任何类型或者形式的输入装置。输入装置 628 的实例包括而不限于：一种键盘、一种指向装置、一种语音识别装置、或任何其他输入装置。在至少一个实施方案中，输入装置 628 可以单独地或与其他元件相结合来执行和 / 或作为一种手段来执行在此披露的确定、计算、识别、应用、使用、询问、接收、和 / 或者发送步骤中的一个或多个。输入装置 628 还可以用于执行和 / 或作为一种手段来执行本披露中提出的其他步骤和特征。

[0061] 如图 6 所示，示例性的计算系统 610 还可以包括通过一个存储接口 634 联接到通信基础结构 612 上的一个主存储装置 632 以及一个后备存储装置 633。存储装置 632 和 633 总体上代表能够存储数据和 / 或其他计算机可读指令的任何类型或形式的存储装置或媒质。例如，存储装置 632 与 633 可以是一种磁盘驱动器（例如，一种所谓的硬盘驱动器）、一种软盘驱动器、一种磁带驱动器、一种光盘驱动器、一种闪存驱动器、或者类似装置。存储接口 634 总体上代表用于在存储装置 632 和 633 与计算系统 610 的其他部件之间传送数据的任何类型或形式的接口或装置。

[0062] 在某些实施方案中，存储装置 632 和 633 可以被配置为用于读取自和 / 或写入到一个可装卸的存储单元，该可装卸的存储单元被配置为用于存储计算机软件、数据、或其他计算机可读信息。合适的可装卸存储单元的实例包括而不限于：一种软盘、一种磁带、一种光盘、一种闪存装置，或诸如此类。存储装置 632 和 633 还可以包括允许将计算机软件、数据、或其他计算机可读指令载入计算系统 610 的其他类似的结构或装置。例如，存储装置 632 和 633 可以被配置用于读取和写入软件、数据、或其他计算机可读信息。存储装置 632 和 633 还可以作为计算系统 610 的一部分或可以是通过其他接口系统访问的一个分离的装置。

[0063] 在某些实施方案中，在此披露的该示例性文件系统可存储在主存储装置 632 中，而在此披露的文件系统的备份可存储在后备存储装置 633 中。例如，存储装置 632 和 633 还可以被用来单独地或与其他元件相结合执行和 / 或作为一种手段来执行在此披露的确定、计算、识别、应用、使用、询问、接收、和 / 或发送步骤中的一项或多项。存储装置 632 和 633 还可以被用于执行和 / 或作为一种手段用于执行本披露提出的其他步骤和特征。

[0064] 很多其他装置或子系统可以被连接到计算系统 610 上。相反，图 6 中所示的所有部件和装置不必都存在以实现在此所说明和 / 或示出的实施方案。以上提及的装置和子系统还能够以不同于图 6 所示的多种方式进行互联。计算系统 610 还可以使用任意数目的软

件、固件和/或硬件配置。例如，在此披露的一个或多个示例性实施方案可以作为一种计算机可读媒质上的计算机程序（还可称为计算机软件、软件应用程序、计算机可读指令、或计算机控制逻辑）进行编码。短语“计算机可读媒质”总体上是指能够存储或携带计算机可读指令的任何形式的装置、载体、或媒质。计算机可读媒质的实例包括而不限于，传输型媒介，如载波，以及物理媒质，如磁性存储媒质（例如硬盘驱动器和软盘驱动器）、光存储媒质（例如 CD-ROM 或 DVD-ROM）、电子存储媒质（例如固态驱动器和闪存媒质），以及其他分布式系统。

[0065] 包括计算机程序的计算机可读媒质可以被载入计算系统 610 中。然后在计算机可读媒质上存储的全部或一部分计算机程序可以被存储到系统内存 616 和/或存储装置 632 和 633 的不同部分中。当由处理器 614 执行时，载入到计算系统 610 中的一个计算机程序可以致使处理器 614 执行和/或作为一种手段用于执行在此所说明和/或展示的一个或多个示例性实施方案的多种功能。额外地或可替代地，在此所说明和/或展示的一个或多个示例性实施方案可以在固件和/或硬件中实施。例如计算系统 610 可被配置用作一种专用集成电路 (ASIC)，它被适配为用于实施在此所说明的一个或多个示例性实施方案。

[0066] 图 7 是一个示例性的网络体系结构 700 的框图，其中，客户系统 710、720、以及 730 与服务器 740 和 745 可被联接到一个网络 750 上。客户系统 710、720、和 730 总体上代表任何类型或形式的计算装置或系统，如图 6 中的示例性计算系统 610。类似地，服务器 740 和 745 总体上代表被配置为用于提供不同的数据库服务和/或运行某种软件应用程序的计算装置或系统，如应用服务器或数据库服务器。网络 750 总体上代表任何电信或计算机网络；例如，它包括：一种内部网、一种广域网 (WAN)、一种局域网 (LAN)、一种个人区域网 (PAN)、或互联网。

[0067] 如图 7 所示，一个或多个存储装置 760(1)-(N) 可直接附接到服务器 740 上。类似地，一个或多个存储装置 770(1)-(N) 可直接附接到服务器 745 上。存储装置 760(1)-(N) 和存储装置 770(1)-(N) 总体上代表能够存储数据和/或其他计算机可读指令的任何类型或形式的存储装置或媒质。在某些实施方案中，存储装置 760(1)-(N) 和存储装置 770(1)-(N) 可代表被配置为用于使用不同协议（例如 NFS、SMB、或 CIFS）来与服务器 740 和 745 进行通信的网络附联存储 (NAS) 装置。

[0068] 服务器 740 和 745 还可以被连接到一种存储器区域网络 (SAN) 光纤通道 780 上。SAN 光纤通道 780 总体上代表能够协助多个存储装置之间互相通信的任何类型或形式的计算机网络或体系结构。SAN 光纤通道 780 可以协助在服务器 740 和 745 与多个存储装置 790(1)-(N) 和/或一个智能存储阵列 795 之间的通信。SAN 光纤通道 780 还可以通过网络 750 以及服务器 740 和 745 协助客户系统 710、720、和 730 以及存储装置 790(1)-(N) 和/或智能存储阵列 795 之间的通信，其方式为使得装置 790(1)-(N) 以及阵列 795 对于客户系统 710、720、和 730 而言表现为好像是本地附联的装置。如同存储装置 760(1)-(N) 和存储装置 770(1)-(N)，存储装置 790(1)-(N) 和智能存储阵列 795 总体上代表能够存储数据和/或其他计算机可读指令的任何类型或形式的存储装置或媒质。

[0069] 在没某些实施方案中，并参照图 6 的示例性计算系统 610，一个通信接口（如图 6 中的通信接口 622）可以被用于在每个客户系统 710、720、和 730 以及网络 750 之间提供连接性。例如，客户系统 710、720、和 730 通过一个网络浏览器或其他客户软件可以能够访问

服务器 740 或 745 上的信息。此类软件可以允许客户系统 710、720、和 730 访问由服务器 740、服务器 745、存储装置 760(1)-(N)、存储装置 770(1)-(N)、存储装置 790(1)-(N)、或智能存储阵列 795 托管的数据。虽然图 7 描绘了使用一个网络（如互联网）用于交换数据，在此说明和 / 或展示的这些实施方案并不局限于互联网或任何特定的基于网络的环境。

[0070] 在至少一个实施方案中，在此披露的一个或多个示例性实施方案的全部或一部分可被编码为一种计算机程序并且由服务器 740、服务器 745、存储装置 760(1)-(N)、存储装置 770(1)-(N)、存储装置 790(1)-(N)、智能存储阵列 795、或它们中的任意组合载入并执行。在此披露的一个或多个示例性实施方案的全部或一部分还可以被编码成为一种计算机程序，它存储在服务器 740 中，由服务器 745 来运行，并在网络 750 上分发给客户系统 710、720、和 730。因此，网络体系结构 700 可以单独地或与其他元件相结合执行和 / 或作为一种手段来执行在此说明的确定、计算、识别、应用、使用、询问、接收、和 / 或者发送步骤中的一个或多个。网络体系结构 700 还可以被用于执行和 / 或作为一种手段用于执行本披露中提出的其他步骤和特征。

[0071] 如以上所详述，计算系统 610 和 / 或网络体系结构 700 的一个或者多个部件可以执行和 / 或作为一种手段用于执行（单独地或者与其他元件相结合）在此说明和 / 或展示的这些示例性实施方案的一个或者多个步骤。例如，一种用于确定文件分类的计算机实施的方法可以包括确定存储在一个第一计算系统上的一个第一文件的信息。该方法还包括通过将该第一文件的识别信息发送给一个第二计算系统来向该第二计算系统询问分类信息。该方法还可以包括接收一个第二文件的识别信息并且接收该分类信息。分类信息可以表明该第一文件和第二文件是可信任的。该方法可以包括使用该第二文件的识别信息来确定该第二文件是存储在该第一计算系统上。该方法可以进一步包括通过将该第一和第二文件从安全扫描中排除来将该分类信息应用于该第一和第二文件。

[0072] 在一些实施方案中，确定该第一文件的识别信息可以包括计算该第一文件的一个摘要。在至少一个实施方案中，该第一文件的识别信息可以包括该第一文件的摘要，并且该第一文件的识别信息可以包括存储该第一文件的一个目录的名称。

[0073] 根据不同的实施方案，该摘要可以包括该第一文件的一个散列并且存储该第一文件的目录的名称可以包括一个标准化的目录路径。根据不同的实施方案，该第二文件的识别信息可以包括该第二文件的一个文件名以及该第二文件的摘要。使用该第二文件的识别信息来确定该第二文件是存储在该第一计算装置上可以包括确定该第一计算装置包括具有该第二文件的文件名的一个文件以及确定该第二文件的摘要与具有该文件名的文件的摘要相匹配。

[0074] 在不同的实施方案中，该第二文件的识别信息包括该第二文件的大小。使用该第二文件的识别信息来确定该第二文件是存储在该第一计算装置上可以包括确定具有该第二文件的文件名的文件的大小与该第二文件的大小相同。

[0075] 根据不同的实施方案，该第一文件被一个第一软件程序安装在该第一计算装置上并且该第二文件也已被该第一软件程序安装在该第一计算装置上。在至少一个实施方案中，该方法可以包括接收对于一组文件中的每个文件的识别信息。该组文件是与安装了该第一文件的一个应用程序相关联的。该组文件可以包括一个第二文件。该方法还可以包括接收对于该组文件中的每个文件的分类信息。根据某些实施方案，该方法可以进一步包括

确定一个文件子组中的每个文件的识别信息。该文件子组可以包括该第一文件。该文件子组是与该应用程序相关联的该组文件的一个子组。该方法还可以包括通过将该文件子组中的每个文件的识别信息发送给该第二计算系统来向该第二计算系统询问分类信息。

[0076] 在某些实施方案中，一种用于确定文件分类的计算机实施的方法可以包括接收存储在该第一计算系统上的一个第一文件的识别信息。该方法还可以包括使用该识别信息在一个分类数据库中搜索对于该第一文件的分类信息。该分类数据库可以存储在一个第二计算系统上。该方法可以包括识别与该第一文件相关联的一个第二文件并且将该第二文件的识别信息发送给该第一计算系统。该方法可以进一步包括将对于该第一文件的分类信息以及对于该第二文件的分类信息发送给该第一计算系统。

[0077] 在一些实施方案中，该第一文件的识别信息包括该第一文件的一个摘要。该第一文件的识别信息还可以包括存储该第一文件的目录的一个名称。根据不同的实施方案，该摘要可以包括该第一文件的一个散列。存储该第一文件的目录的名称可以包括一个标准化的目录路径。

[0078] 在一些实施方案中，该第一文件可以被一个第一软件程序安装在该第一计算装置上。在不同的实施方案中，识别该第二文件可以包括使用存储该第一文件的目录的名称来发现与该第一软件程序相关联的一个目录。在一些实施方案中，识别该第二文件可以包括确定与该第一软件程序相关联的目录可能包括该第二文件。

[0079] 根据至少一个实施方案，该第二文件的识别信息可以包括该第二文件的多个文件名、该第二文件的大小、以及该第二文件的一个摘要中的至少一项。该方法还可以包括发送对于一组文件中的每个文件的识别信息。该组文件可以是与安装了该第一文件的一个应用程序相关联的。该方法还可以包括接收对于该组文件中的每个文件的分类信息。根据不同的实施方案，该方法可以进一步包括接收一个文件子组中的每个文件的识别信息。可以从该第一计算系统接收该识别信息。该文件子组可以包括该第一文件，并且该文件子组可以是与该应用程序相关联的该组文件的一个子组。

[0080] 虽然以上披露使用了多个具体的框图、流程图、以及实例阐明了不同的实施方案，在此说明和/或展示的每个框图部件、流程图步骤、操作、和/或部件都可以单独地和/或共同地使用一个大范围的硬件、软件、或者固件（或者它们的任何组合）配置来实施。另外，在其他部件之中所包括的任何部件的披露都应该看作本质上是示例性的，因为可以实施许多其他的体系结构来达到同样的功能。

[0081] 在此说明和/或展示的进程的参数以及步骤的顺序仅仅是以举例的方式给出并且可以按希望来更改。例如，虽然在此展示和/或说明的这些步骤可以按照一个具体的顺序来示出或讨论，但这些步骤并非必须按照所展示或者所讨论的顺序来执行。在此说明和/或展示的不同的示例性方法还可以省略在此说明或展示的一个或者多个步骤或者还可以包括除所披露的那些之外的额外步骤。

[0082] 此外，虽然不同的实施方案在此已经在全功能性计算系统的背景下进行了说明和/或展示，这些示例性实施方案中的一个或者多个能够以多种形式作为一个程序产品来分发，不管实际用于进行该分发的计算机可读媒介的具体形式如何。在此披露的这些实施方案还可以通过使用执行一些特定任务的软件模块来实施。这些软件模块可以包括脚本、成批文件、或者其他可执行文件，它们可以存储在一种计算机可读的存储媒质上或者在一种

计算系统中。在一些实施方案中，这些软件模块可以将一个计算系统配置用于实施在此披露的一个或者多个示例性的实施方案。

[0083] 已经提供了以上说明用于使本领域的其他普通技术人员能够最好地使用在此披露的这些示例性实施方案的不同方面。这种示例性说明并非旨在是穷尽性的或者被限制在所披露的任何准确的形式上。许多修改与变更都是可能的而不背离本披露的精神与范围。应该认为在此披露的这些实施方案在所有方面都是展示性的而非限制性的。应该参照所附权利要求及其等效物来确定本披露的范围。

[0084] 除非另外说明，如在本说明书与权利要求中所使用的，术语“一种”或“一个”将被解释为“至少一个”的意思。此外，为便于使用，如在本说明书以及权利要求中所使用的文字“包含”和“具有”是可以互换的并且具有与文字“包括”相同的含义。

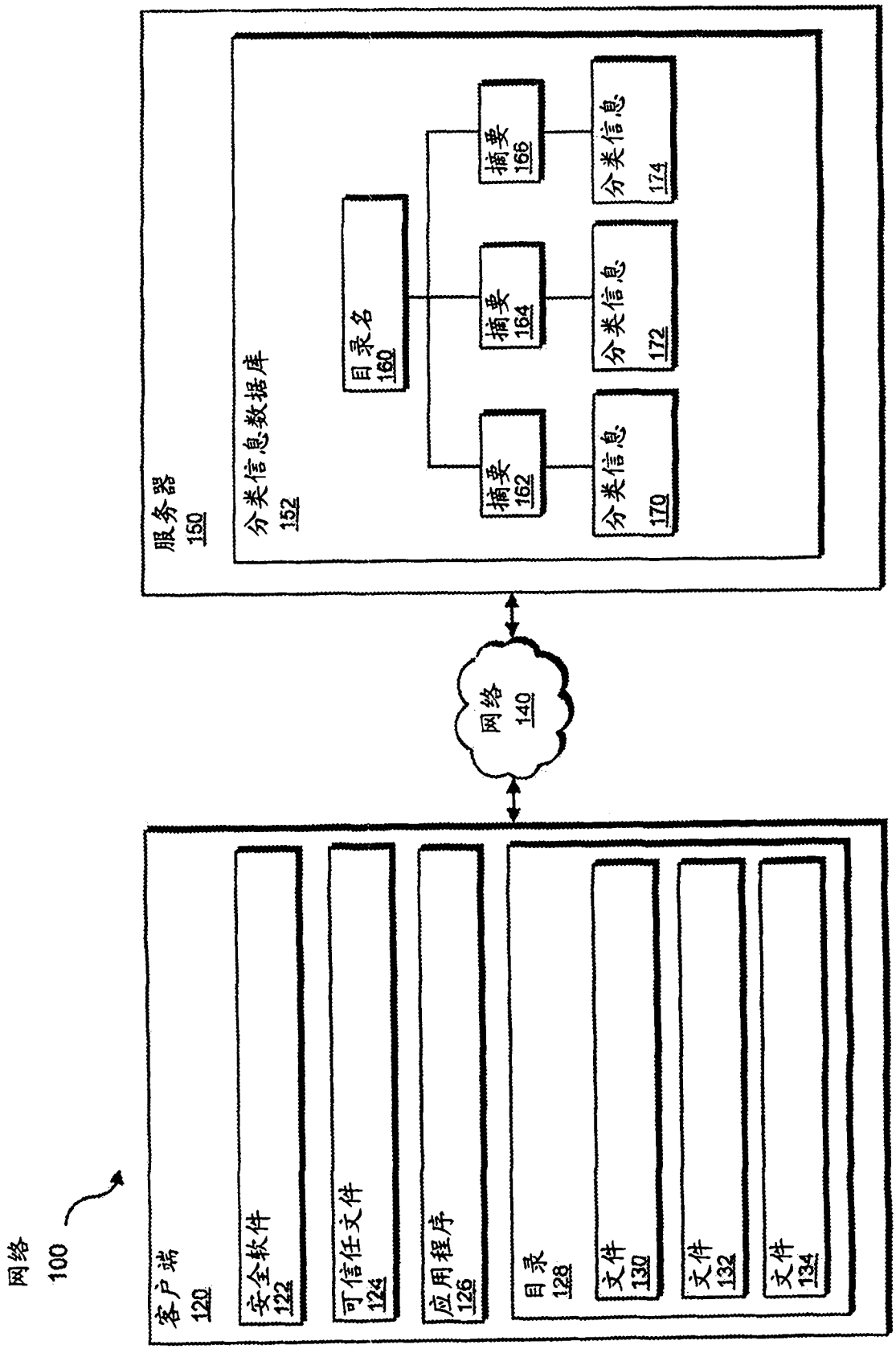

网络
100
客户端
120
安全软件
122
可信任文件
124
应用程序
126
目录
128
文件
130
文件
132
文件
134
网络
140
服务器
150
分类信息数据库
152
目录名
160
摘要
162
摘要
164
摘要
166
分类信息
170
分类信息
172
分类信息
174

图 1

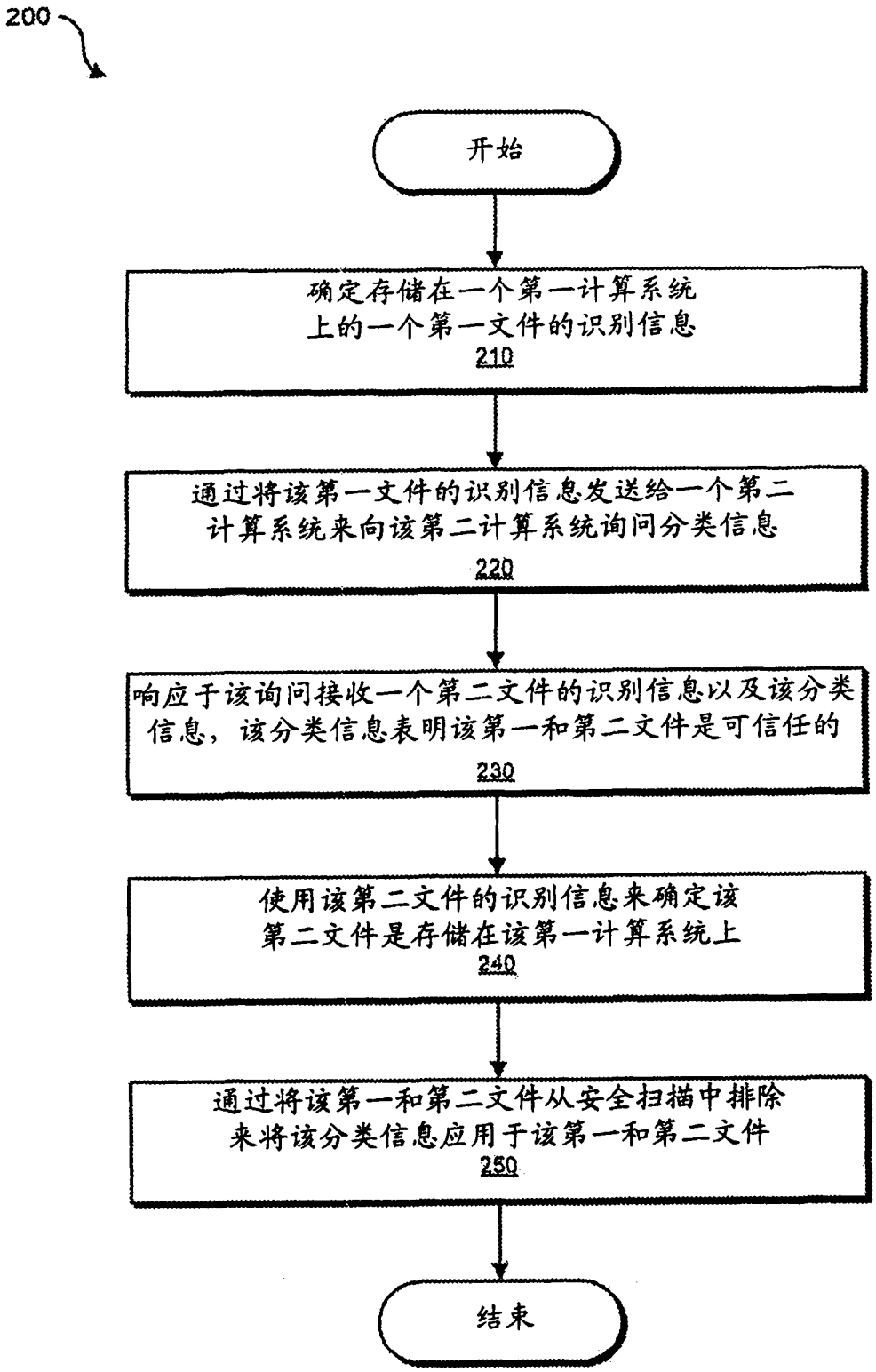

200
开始
确定存储在一个第一计算系统上的一个第一文件的识别信息
210
通过将该第一文件的识别信息发送给一个第二计算系统来向该第二计算系统询问分类信息
220
响应于该询问接收一个第二文件的识别信息以及该分类信息，该分类信息表明该第一和第二文件是可信任的
230
使用该第二文件的识别信息来确定该第二文件是存储在该第一计算系统上
240
通过将该第一和第二文件从安全扫描中排除来将该分类信息应用于该第一和第二文件
250
结束

图 2

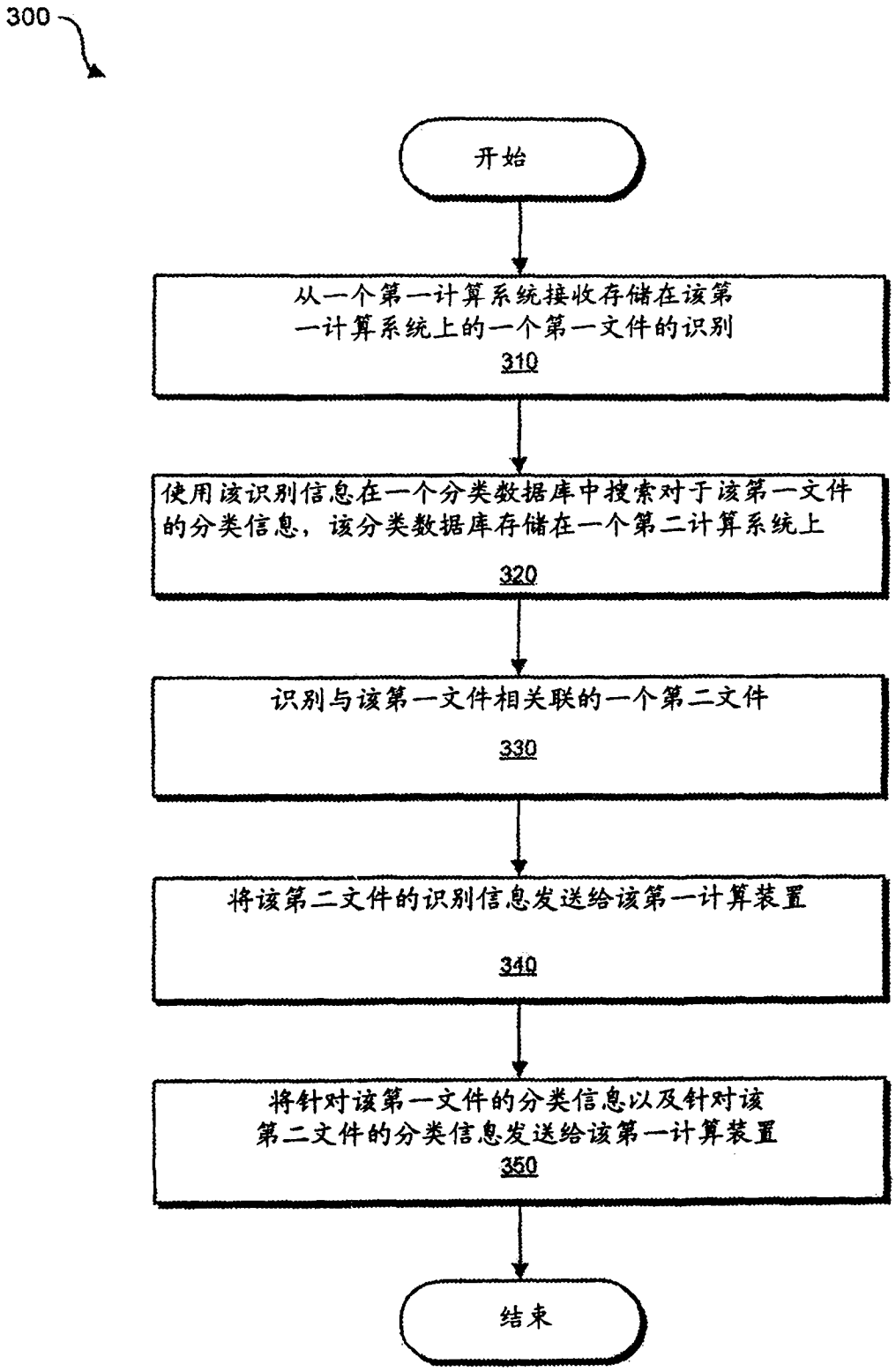

300
开始
从一个第一计算系统接收存储在该第一计算系统上的一个第一文件的识别
310
使用该识别信息在一个分类数据库中搜索对于该第一文件的分类信息，该分类数据库存储在一个第二计算系统上
320
识别与该第一文件相关联的一个第二文件
330
将该第二文件的识别信息发送给该第一计算装置
340
将针对该第一文件的分类信息以及针对该第二文件的分类信息发送给该第一计算装置
350
结束

图 3

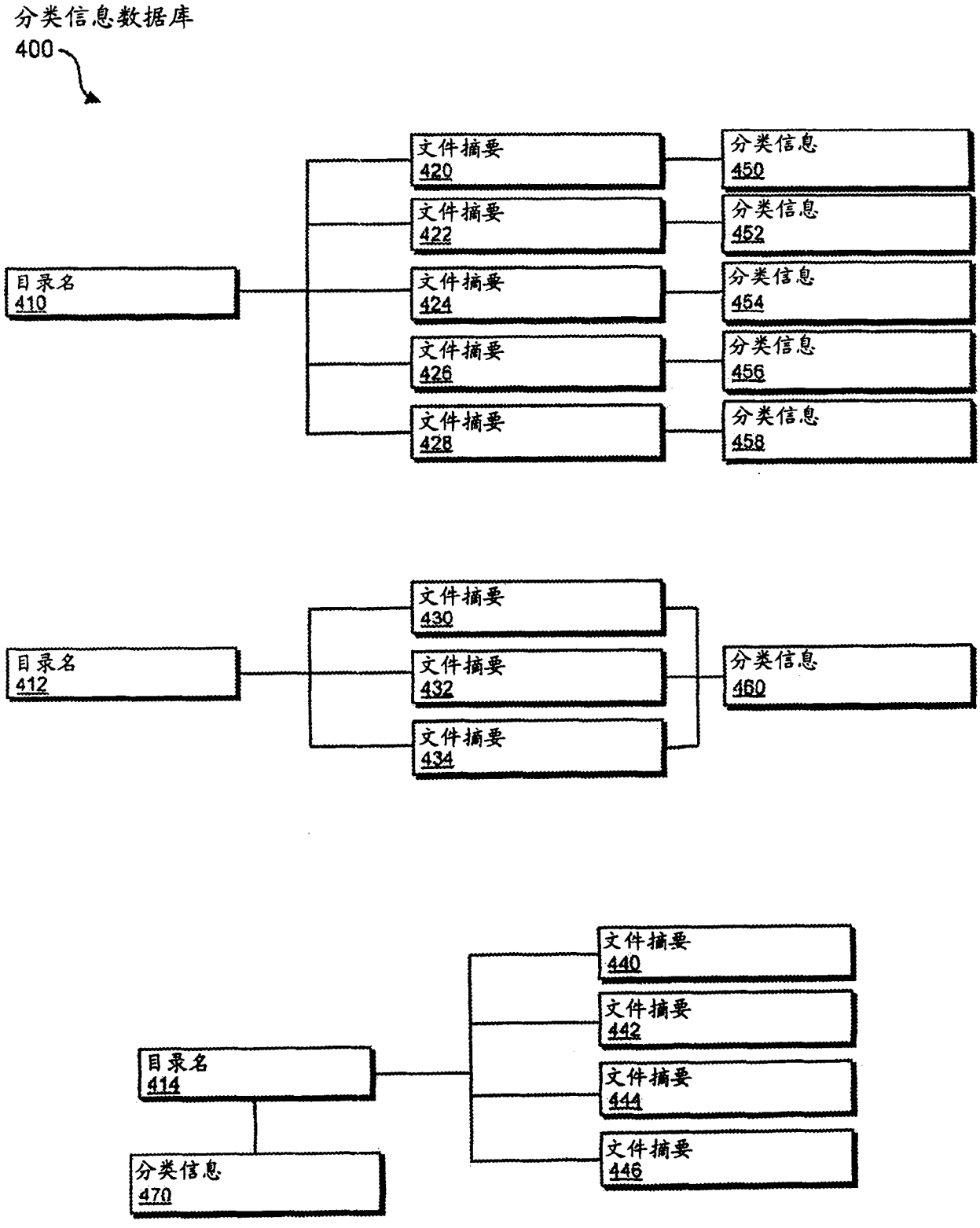

分类信息数据库
400
目录名
410
文件摘要
420
文件摘要
422
文件摘要
424
文件摘要
426
文件摘要
428
分类信息
450
分类信息
452
分类信息
454
分类信息
456
分类信息
458
目录名
412
文件摘要
430
文件摘要
432
文件摘要
434
分类信息
460
目录名
414
分类信息
470
文件摘要
440
文件摘要
442
文件摘要
444
文件摘要
446

图 4

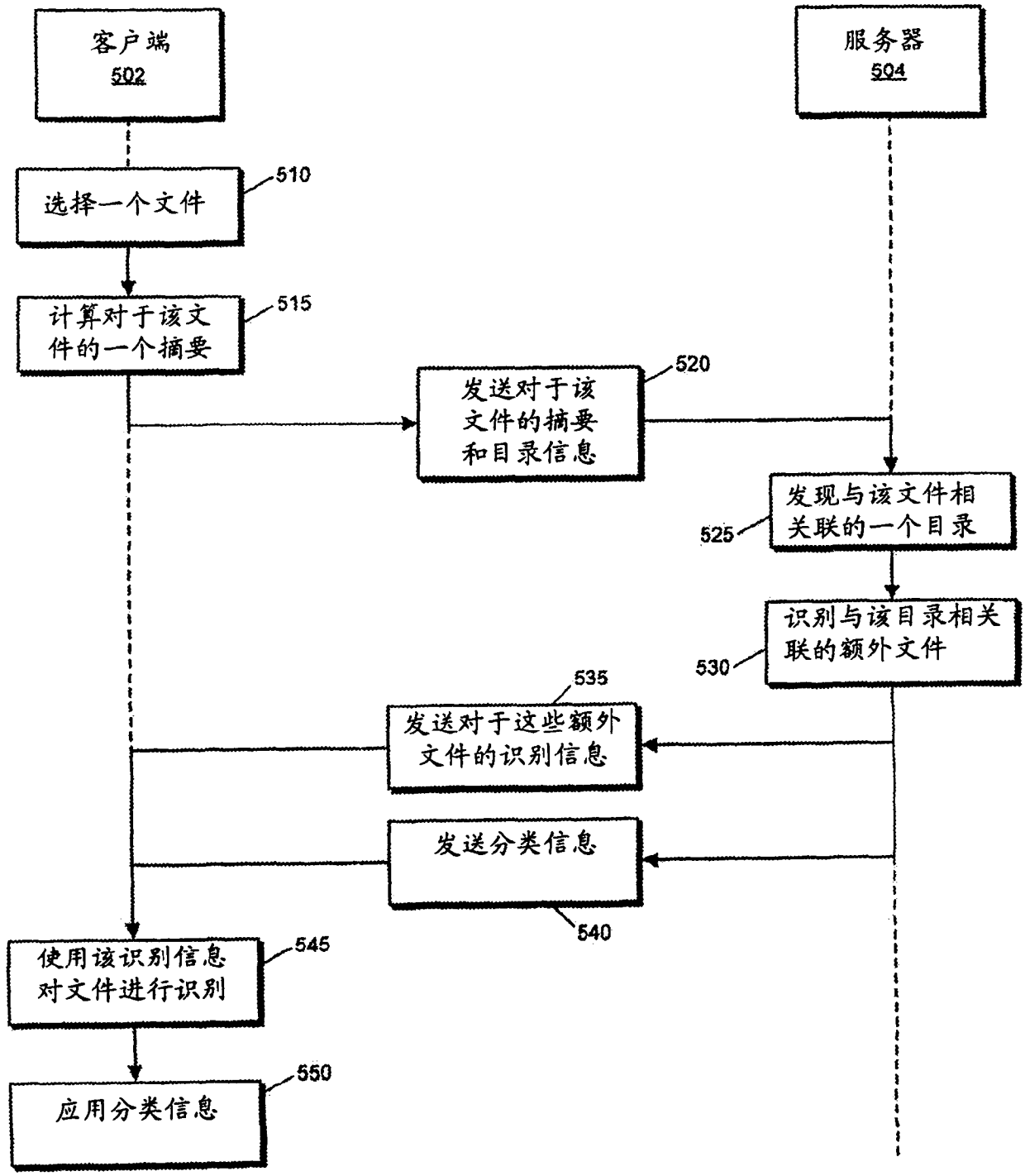

客户端
502
服务器
504
选择一个文件
510
计算对于该文件的一个摘要
515
发送对于该文件的摘要和目录信息
520
发现与该文件相关联的一个目录
525
识别与该目录相关联的额外文件
530
发送对于这些额外文件的识别信息
535
发送分类信息
540
使用该识别信息对文件进行识别
545
应用分类信息
550

图 5

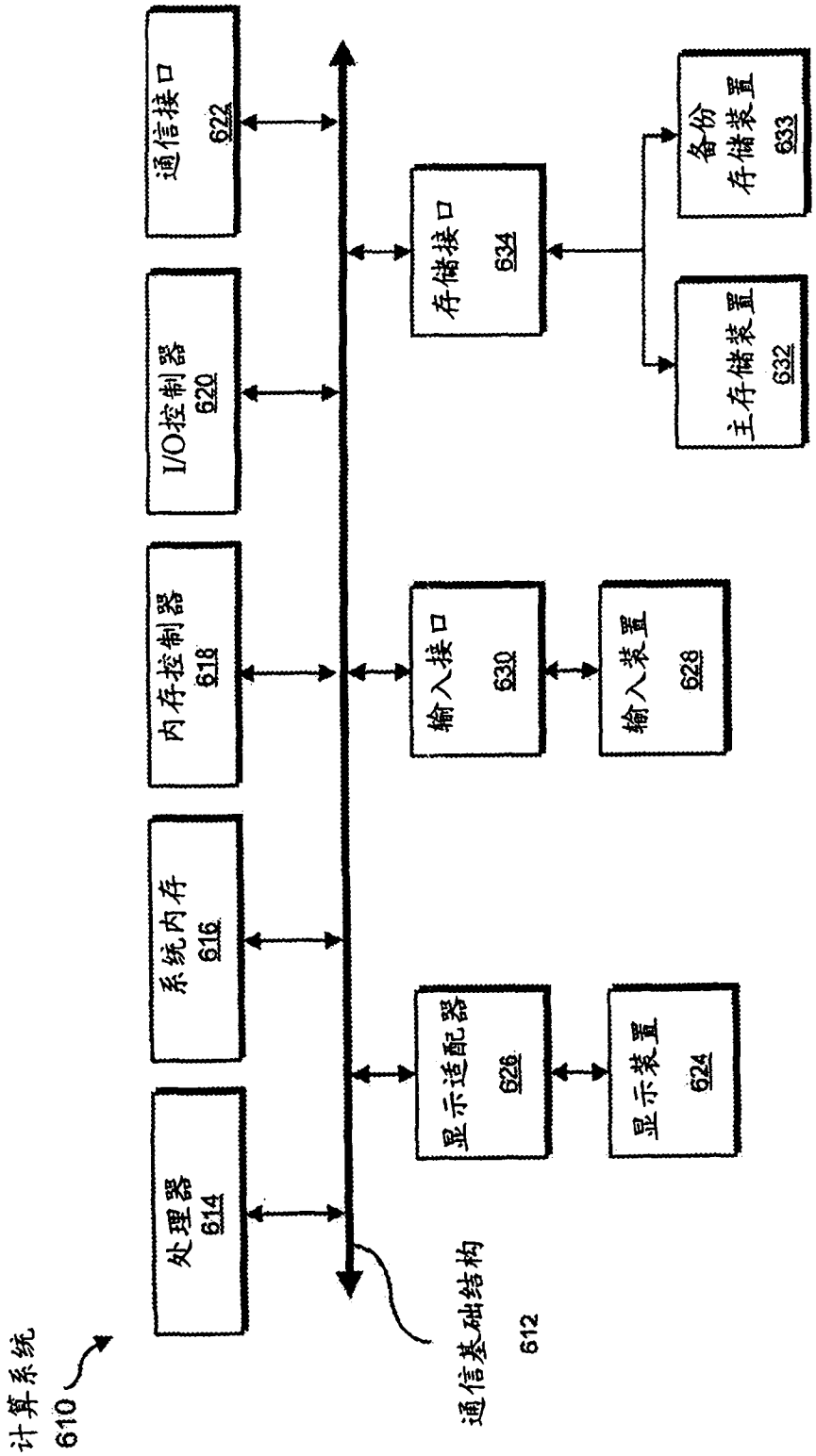

计算系统
610
处理器
614
系统内存
616
内存控制器
618
I/O控制器
620
通信接口
622
通信基础结构
612
显示适配器
626
显示装置
624
输入接口
630
输入装置
628
存储接口
634
主存储装置
632
备份
存储装置
633

图 6

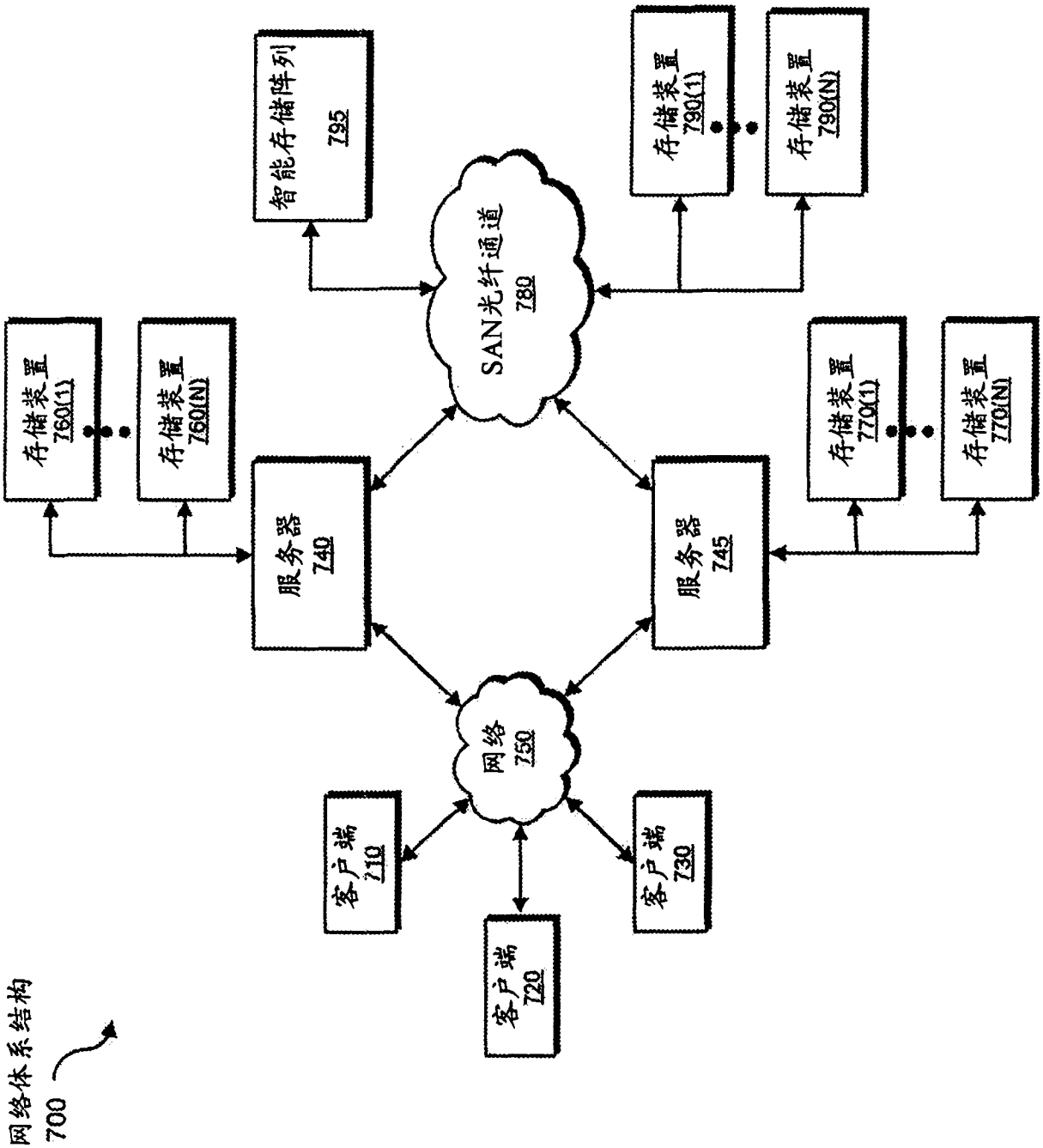

网络体系结构
700
客户端
710
客户端
720
客户端
730
网络
750
服务器
740
服务器
745
存储装置
760(1)
存储装置
760(N)
SAN光纤通道
780
智能存储阵列
795
存储装置
790(1)
存储装置
790(N)
存储装置
770(1)
存储装置
770(N)

图 7