(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112149122 B
(45)授权公告日 2025.05.30

(21)申请号 202010222091.9

(22)申请日 2020.03.26

(65)同一申请的已公布的文献号
申请公布号 CN 112149122 A

(43)申请公布日 2020.12.29

(30)优先权数据
2019120230 2019.06.28 RU
16/654,434 2019.10.16 US

(73)专利权人 卡巴斯基实验室股份制公司
地址 俄罗斯联邦莫斯科列宁格勒斯科公路39A3

(72)发明人 弗拉基米尔·A·库斯科夫
尼基塔·A·布卡 安东·A·基瓦
奥列格·P·沃尔科夫
德米特里·Y·卢卡舍维奇
叶夫根尼·A·罗金斯基
康斯坦丁·M·菲拉托夫
德米特里·V·拉托欣

(74)专利代理机构 北京高沃律师事务所 11569
专利代理师 刘凤玲

(51)Int.Cl.
G06F 18/2431(2023.01)
G06F 18/2415(2023.01)
G06F 21/56(2013.01)

(56)对比文件
US 2012240234 A1,2012.09.20

审查员 薛慧

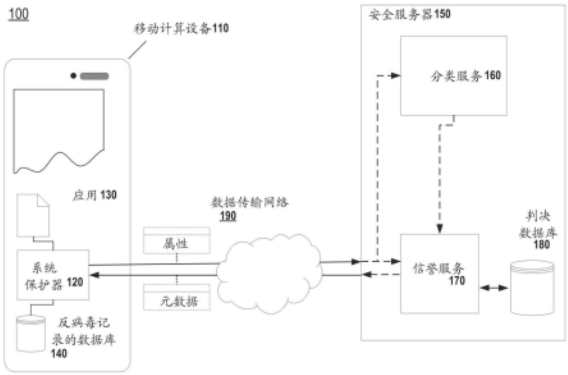
权利要求书3页 说明书12页 附图4页

(54)发明名称

对计算设备上的应用进行分类的系统和方法

(57)摘要

本文公开了用于对计算设备上的应用进行分类的系统和方法。在一方面,示例性方法包括,从安全服务器获取应用的分类的结果,当分类的结果满足相关性规则时,将分类的结果指定为相关的,并且基于结果被指定为相关的来确定应用的类别,并且当分类的结果不满足相关性规则时,执行以下中的至少一个:终止应用的分类,并且基于应用的属性集来更新应用的分类。



CN 112149122 B

1.一种用于对计算设备上的应用进行分类的方法,所述方法包括:

从安全服务器获取应用的分类的结果,其中,所述应用的分类的结果被呈现为所述应用属于应用的一个或更多个相应类别的一个或更多个概率,所述应用的一个或更多个类别包括:恶意应用的类别、既不是恶意也不是可信的不需要应用的类别和可信应用的类别;

将一个或更多个相关性规则应用于所述分类的结果,其中所述一个或更多个相关性规则包含对所述分类的结果的要求,所述要求降低应用的分类中的假阳性错误的概率,其中,与所述分类的结果相关的一个或更多个相关性规则基于以下条件中的至少一个:分类的时间戳和当前时间戳之间的差在允许值的指定范围内,以及专家分类系统的错误估计不大于预设阈值;

当所述分类的结果满足一个或更多个相关性规则时,将所述分类的结果指定为相关的,并且基于所述结果被指定为相关的来确定所述应用的类别;以及

当所述分类的结果不满足所述相关性规则时,执行以下中的至少一个:终止所述应用的分类,并且基于所述应用的属性集来更新所述应用的分类,其中,所述分类的更新包括:响应于将所述属性集发送至安全服务器,从所述安全服务器接收所更新的分类。

2.根据权利要求1所述的方法,其中,所述预设阈值为0.5%。

3.根据权利要求1所述的方法,其中,所述应用的属性集包括以下中的至少一个:

所述应用的应用包中的多个文件;

所述应用包中的多个可执行文件;

多个请求的许可以及许可的类型;

所述应用包中的所述可执行文件中的多个类;以及

所述应用包中的所述可执行文件中的多个方法。

4.根据权利要求1所述的方法,其中,使用至少部分基于所述分类的结果的启发式规则来确定所述应用的类别。

5.根据权利要求1所述的方法,还包括:

当所述应用被分类为恶意时,从所述计算设备移除所述应用或者隔离所述应用;以及

当所述应用被分类为不需要时,执行以下中的至少一个:从所述计算设备移除所述应用,向所述计算设备的用户通知所述不需要应用在所述计算设备上的存在,向所述计算设备的所述用户提供用于选择保留或移除所述不需要应用的选项,以及撤销先前授予所述应用的许可。

6.根据权利要求1所述的方法,其中,如果所述分类的结果满足至少一个相关性规则,则所述应用的分类的结果被指定为满足相关性规则。

7.根据权利要求1所述的方法,其中,如果所述分类的结果满足所有相关性规则,则所述应用的分类的结果被指定为满足相关性规则。

8.根据权利要求1所述的方法,其中,所述可信应用的类别包括由可信软件制造商开发的应用、从可信来源下载的应用,或者其标识符保存在可信应用的数据库中的应用。

9.一种用于对计算设备上的应用进行分类的系统,包括:

至少一个处理器,被配置为:

从安全服务器获取应用的分类的结果,其中,所述应用的分类的结果被呈现为所述应用属于应用的一个或更多个相应类别的一个或更多个概率,所述应用的一个或更多个类别

包括:恶意应用的类别、既不是恶意也不是可信的不需要应用的类别和可信应用的类别;

将一个或更多个相关性规则应用于所述分类的结果,其中所述一个或更多个相关性规则包含对所述分类的结果的要求,所述要求降低应用的分类中的假阳性错误的概率,其中,与所述分类的结果相关的一个或更多个相关性规则基于以下条件中的至少一个:分类的时间戳和当前时间戳之间的差在允许值的指定范围内,以及专家分类系统的错误估计不大于预设阈值;

当所述分类的结果满足一个或更多个相关性规则时,将所述分类的结果指定为相关的,并且基于所述结果被指定为相关的来确定所述应用的类别;以及

当所述分类的结果不满足所述相关性规则时,执行以下中的至少一个:终止所述应用的分类,并且基于所述应用的属性集来更新所述应用的分类,其中,所述分类的更新包括:响应于将所述属性集发送至安全服务器,从所述安全服务器接收所更新的分类。

10.根据权利要求9所述的系统,其中,所述预设阈值为0.5%。

11.根据权利要求9所述的系统,其中,所述应用的所述属性集包括以下中的至少一个:

所述应用的应用包中的多个文件;

所述应用包中的多个可执行文件;

多个请求的许可和许可的类型;

所述应用包中的所述可执行文件中的多个类;以及

所述应用包中的所述可执行文件中的多个方法。

12.根据权利要求9所述的系统,其中,使用至少部分基于所述分类的结果的启发式规则来确定所述应用的类别。

13.根据权利要求9所述的系统,所述处理器还被配置为:

当所述应用被分类为恶意时,从所述计算设备移除所述应用或者隔离所述应用;以及

当所述应用被分类为不需要时,执行以下中的至少一个:从所述计算设备移除所述应用,向所述计算设备的用户通知所述不需要应用在所述计算设备上的存在,向所述计算设备的所述用户提供用于选择保留或移除所述不需要应用的选项,以及撤销先前授予所述应用的许可。

14.根据权利要求9所述的系统,其中,如果所述分类的结果满足至少一个相关性规则,则所述应用的分类的结果被指定为满足相关性规则。

15.根据权利要求9所述的系统,其中,如果所述分类的结果满足所有相关性规则,则所述应用的分类的结果被指定为满足相关性规则。

16.根据权利要求9所述的系统,其中,所述可信应用的类别包括由可信软件制造商开发的应用、从可信来源下载的应用,或者其标识符保存在可信应用的数据库中的应用。

17.一种非暂时性计算机可读介质,存储用于对计算设备上的应用进行分类的计算机可执行指令,包括用于以下操作的指令:

从安全服务器获取应用的分类的结果,其中,所述应用的分类的结果被呈现为所述应用属于应用的一个或更多个相应类别的一个或更多个概率,所述应用的一个或更多个类别包括:恶意应用的类别、既不是恶意也不是可信的不需要应用的类别和可信应用的类别;

将一个或更多个相关性规则应用于所述分类的结果,其中所述一个或更多个相关性规则包含对所述分类的结果的要求,所述要求降低应用的分类中的假阳性错误的概率,其中,

与所述分类的结果相关的一个或更多个相关性规则基于以下条件中的至少一个:分类的时间戳和当前时间戳之间的差在允许值的指定范围内,以及专家分类系统的错误估计不大于预设阈值;

当所述分类的结果满足一个或更多个相关性规则时,将所述分类的结果指定为相关的,并且基于所述结果被指定为相关的来确定所述应用的类别;以及

当所述分类的结果不满足所述相关性规则时,执行以下中的至少一个:终止所述应用的分类,并且基于所述应用的属性集来更新所述应用的分类,其中,所述分类的更新包括:响应于将所述属性集发送至安全服务器,从所述安全服务器接收所更新的分类。

18.根据权利要求17所述的非暂时性计算机可读介质,其中,所述预设阈值为0.5%。

19.根据权利要求17所述的非暂时性计算机可读介质,其中,所述应用的所述属性集包括以下中的至少一个:

所述应用的应用包中的多个文件;

所述应用包中的多个可执行文件;

多个请求的许可和许可的类型;

所述应用包中的所述可执行文件中的多个类;以及

所述应用包中的所述可执行文件中的多个方法。

20.根据权利要求17所述的非暂时性计算机可读介质,其中使用至少部分基于所述分类的结果的启发式规则来确定所述应用的类别。

21.根据权利要求17所述的非暂时性计算机可读介质,所述指令还包括用于以下操作的指令:

当所述应用被分类为恶意时,从所述计算设备移除所述应用或者隔离所述应用;以及

当所述应用被分类为不需要时,执行以下中的至少一个:从所述计算设备移除所述应用,向所述计算设备的用户通知所述不需要应用在所述计算设备上的存在,向所述计算设备的所述用户提供用于选择保留或移除所述不需要应用的选项,以及撤销先前授予所述应用的许可。

22.根据权利要求17所述的非暂时性计算机可读介质,其中,如果所述分类的结果满足至少一个相关性规则,则所述应用的分类的结果被指定为满足相关性规则。

23.根据权利要求17所述的非暂时性计算机可读介质,其中,如果所述分类的结果满足所有相关性规则,则所述应用的分类的结果被指定为满足相关性规则。

24.根据权利要求17所述的非暂时性计算机可读介质,其中,所述可信应用的类别包括由可信软件制造商开发的应用、从可信来源下载的应用,或者其标识符保存在可信应用的数据库中的应用。

对计算设备上的应用进行分类的系统和方法

技术领域

[0001] 本公开涉及使用远程服务器检测计算设备上的恶意应用的领域。

背景技术

[0002] 包括移动计算设备在内的计算设备的广泛普及,为罪犯进行例如使用恶意软件进行网络攻击开放了广阔的前景。通过获取对用户的计算设备的非法访问,罪犯能获取对用户的机密数据和对用户的通信的访问。使用非法访问的内容,罪犯能够以用户的名义执行动作,包括进行金融交易的动作。因此,通常使用专用软件(例如,反病毒软件)来保护用户的设备。

[0003] 现代反病毒软件经常与远程基础设施协同工作,每个可能的服务都在该远程基础设施内操作。例如,远程基础设施可以用于提供数据(诸如关于应用是否属于某些类别的数据)的服务。例如,根据反病毒软件的请求,远程基础设施可以提供指示应用是否属于某些类别的数据。这样的服务通过接管诸如图像识别或使用多个标准对对象(文件、应用等)进行分类等劳动密集型计算任务,来减少用户的计算设备上的负担。

[0004] 但是,这种方法具有其缺点。连接到远程服务器的大量反病毒应用对远程基础设施的服务造成重负。附带地,来自安装在各种设备上的反病毒应用的许多请求是相同的。即,大量的请求是获取关于完全相同的应用或文件的类别的信息。为了避免在远程基础设施上为每次接收到的请求运行相同的任务,缓存服务可用于记忆先前执行任务的结果。尽管这种方法在一定程度上减少了计算负担,但结果增加了第一类和第二类错误(假阳性和假阴性)。因此,上述方法具有缺点。

[0005] 因此,需要一种消除对增加计算资源的需要,并且减少在应用的分类中的错误的同时又能检测恶意应用的更优的方式。

发明内容

[0006] 本公开的各方面涉及信息安全领域,尤其涉及用于应用的分类的系统和方法。

[0007] 在一个示例性方面,用于计算设备上的应用的分类的方法在包括硬件处理器的计算机中实施,该方法包括:从安全服务器获取应用的分类结果,当分类结果满足相关性规则时,将分类结果指定为相关,并且基于结果被指定为相关的来确定应用的类别,并且当分类结果不满足相关性规则时,执行以下中的至少一个:终止应用的分类,并且基于应用的属性集来更新应用的分类。

[0008] 根据本公开的一方面,提供一种用于对计算设备上的应用进行分类的系统,该系统包括硬件处理器,该硬件处理器被配置为:从安全服务器获取应用的分类结果,当分类结果满足相关性规则时,将分类结果指定为相关,并且基于结果被指定为相关的来确定应用的类别,并且当分类结果不满足相关性规则时,执行以下中的至少一个:终止应用的分类,并且基于应用的属性集来更新应用的分类。

[0009] 在一个示例性方面,提供非暂时性计算机可读介质,其上存储有用于对计算设备

上的应用进行分类的指令集,其中,指令集包括用于以下操作的指令:从安全服务器获取应用的分类结果,当分类结果满足相关性规则时,将分类结果指定为相关,并且基于结果被指定为相关的来确定应用的类别,并且当分类结果不满足相关性规则时,执行以下中的至少一个:终止应用的分类,并且基于应用的属性集来更新应用的分类。

[0010] 在一方面,分类的更新包括响应于向安全服务器发送属性集而从安全服务器接收更新的分类。

[0011] 在一方面,应用的属性集包括以下中的至少一个:应用的应用包中的多个文件、应用包中的多个可执行文件、多个被请求的许可和许可的类型、应用包中的可执行文件中的多个类、以及应用包中的可执行文件中的多个方法。

[0012] 在一方面,应用的分类结果被呈现为属于应用的一个或更多个相应类别的应用的一个或更多个概率。

[0013] 在一方面,应用的一个或更多个类别包括以下中的至少一个:恶意应用的类别、不需要的应用的类别和可信应用的类别。

[0014] 在一方面,使用启发式规则来确定应用的类别,该启发式规则至少部分基于分类的结果。

[0015] 在一方面,该方法还包括:当应用被分类为恶意时,从计算设备移除该应用或隔离该应用,并且当应用被分类为不需要时,执行以下中的至少一项:从计算设备移除该应用,通知计算设备的用户在计算设备上不需要的应用的存在,向计算设备的用户提供用于选择保留或移除不需要的应用的选项,以及撤销先前授予该应用的许可。

[0016] 在一方面,本公开的方法对应用进行分类,同时消除对增加计算资源的需要,并且同时降低在应用的分类中的错误。该方法的目的是提高计算机安全。因此,本公开的方法有助于实现数据的信息安全。

附图说明

[0017] 并入本说明书并构成本说明书的一部分的附图示出了本公开的一个或更多个示例性方面,并且与详细描述一起用于解释其原理和实施。

[0018] 图1示出了根据本公开的方面的用于对计算设备上的应用进行分类的系统示意图。

[0019] 图2示出了根据本公开的方面的用于使用分类服务来对应用进行分类的方法。

[0020] 图3示出了根据本公开的方面的用于基于相关性规则对计算设备上的应用进行分类的方法。

[0021] 图4呈现了可以在其上实现本公开的各方面的通用计算机系统的示例。

具体实施方式

[0022] 本文在用于对计算设备上的应用进行分类而不增加对计算资源的需要并且不增加在应用的分类中的错误的系统、方法和计算机程序的上下文中描述示例性方面。本领域普通技术人员将认识到,以下描述仅是说明性的,而并不意在以任何方式进行限制。受益于本公开的本领域技术人员将容易想到其他方面。现在将详细描述如附图所示的示例方面的实施方式。在整个附图和以下描述中,相同的参考标记将尽可能地用于指代相同或相似的

事物。

[0023] 为了清楚地呈现本公开的教导,本文定义了在描述本公开的各个方面中使用的多个术语和概念。

[0024] 恶意应用是能够对计算系统或计算系统(也就是计算机、计算机组、个人计算机、服务器、移动电话等)的用户的数据造成损害的应用,诸如:因特网蠕虫、键盘记录器、计算机病毒等。所造成的损害可以是对计算机资源的非法访问,包括用于盗窃目的而保存在计算机上的数据,以及对资源的非法使用,包括用于数据的存储、执行计算等。

[0025] 可信应用是不会对计算系统或计算系统的用户造成损害的应用。可信应用可以包括由可信软件制造商开发的、从可信源(诸如在可信站点的数据库中列出的站点)下载的应用,或者其标识符(或者该应用由其被唯一识别的其它数据,例如应用的文件的散列值)被保存在可信应用的数据库中的应用。制造商的诸如数字证书等标识符,也可以保存在可信应用的数据库中。

[0026] 不需要的应用既不是恶意的也不是可信的应用。此外,这样的应用能够执行对计算机资源(包括保存在计算机上的数据)的非法访问,虽然这样的数据的泄露不会对计算机或计算机的用户造成直接损害。不需要的应用的示例可以是广告软件,其能够从用户的设备收集数据和/或经由设备向用户显示广告资料。

[0027] 不可信应用既不是可信的也不是不需要的应用,而是例如在反病毒应用的帮助下未被分类为有害的应用。此外,例如,借助于反病毒扫描,不可信任应用随后可以被分类为恶意的。

[0028] 恶意文件是作为恶意应用的组件的文件,并且包含程序代码(例如,可执行或解释代码)。

[0029] 不可信文件是作为不可信应用的组件的文件,并且包含程序代码(例如,可执行或解释代码)。

[0030] 可信文件是作为可信应用的组件的文件。

[0031] 不需要的文件是作为不需要的应用的组件的文件,并且包含程序代码(可执行或可解释代码)。

[0032] 应用的类别是应用的特性,其定义了应用与以下之一的从属关系:可信应用的类别(应用是可信的)、恶意应用的类别(应用是恶意的)或不需要的应用的类别(应用是不想要的)。

[0033] 相关性规则是包含分类结果的要求的规则,其中满足这些要求降低了分类结果中第一类和第二类错误的概率(并且因此也降低了错误的数量)。概率被计算为数值。

[0034] 应用的分类结果是应用属于给定应用的类别的概率。

[0035] 在一方面,本公开描述了一种用于对计算设备上的应用进行分类而不增加对计算资源的需要并且不增加在计算系统(例如,服务器、计算机等)上实施的应用的分类中的错误的系统,该系统包括借助于诸如集成微电路(专用集成电路,ASIC)或现场可编程门阵列(FPGA)等硬件实现的或者例如以诸如微处理器系统和程序指令集等软件和硬件相结合的形式实现的,以及还在神经突触芯片上实现的真实世界的设备、系统、组件和组件组。系统的这种装置的功能可以单独地通过硬件来实现,并且也可以以组合的形式来实现,其中系统装置的一些功能通过软件来实现,一些通过硬件来实现。在某些方面,组件、系统等的一

些或全部可以在通用计算机(诸如图4中所示的通用计算机)的处理器上执行。此外,系统组件可以在单个计算设备内实现,或者分布在若干个互连的计算设备中实现。

[0036] 图1示出了根据本公开的方面的用于对计算设备上的应用进行分类的系统100的示意图。

[0037] 用于对应用进行分类的系统100包括在系统保护器120和安全服务器150上实施的分类服务160。在一方面,系统保护器120被实施在用户的计算设备上,例如在用户的移动计算设备110上实施。在不失一般性的情况下,术语"移动计算设备"用于描述本公开的教导。换句话说,该方法可以被部署在任何标准计算设备上,并且术语"移动计算设备"的使用不意在将本公开的益处仅限制于移动设备。相反,任何计算设备的用户,移动的或其他的,可以从本公开的教导中受益。

[0038] 在一方面,安全服务器150还包括信誉服务170。此外,信誉服务170可以通信地耦接到判决数据库180,该判决数据库180也可以被在安全服务器150上实施。应当注意,安全服务器150可以被作为单个计算设备呈现或作为例如通过数据传输网络190链接的若干互连的计算设备呈现,每个计算设备可以是物理的或虚拟计算设备。

[0039] 在一方面,移动计算设备110还包括反病毒记录的数据库140。在一方面,系统保护器120通信地耦接到反病毒记录的数据库140。例如存储在数据库140中的反病毒记录包括通过反病毒软件(或诸如系统保护器120等类似的系统)为确定应用的类别使用的形式化数据集,诸如用于检测恶意应用等。

[0040] 在一方面,移动计算设备110包括应用130。在一方面,应用130可以是通过数据传输网络190从因特网下载的应用,例如从应用商店(例如,从App商店、谷歌市场(Google Play)等)下载的应用。在另一方面,应用130可以以其它方式获取,例如,经由可移动存储介质或蓝牙连接。为了确保移动计算设备110的安全,系统保护器120被部署在设备110上。

[0041] 在一方面,系统保护器120收集应用130的属性集。在一方面,可由系统保护器120代表应用130收集的属性集包括:

[0042] • 应用130的应用包中的多个文件;

[0043] • 应用包中的多个可执行文件;

[0044] • 多个被请求的许可和许可的类型;

[0045] • 可执行文件中的多个类;以及

[0046] • 可执行文件中的多个方法。

[0047] 在一方面,应用130的应用包可以包括用于应用130的文件的容器。例如,应用包可以包括安装包(APK)档案文件、压缩(ZIP)档案文件或基于任何其它标准的容器,其中APK档案文件被设计为用于在安卓(Android)操作系统的控制下的移动计算设备上运行应用。

[0048] 在一方面,可执行文件包括DEX文件,其被设计为用于在Android操作系统的控制下的设备的执行。

[0049] 在一方面,可执行文件包括包含用于由计算设备执行(其可以包括解释器的使用)的指令的另外的文件。

[0050] 在一方面,对许可的请求是应用130在移动计算设备110上执行某些动作的指示,其中,对于正在请求许可的动作需要移动计算设备110的用户的明确同意。需要明确许可的动作的一些示例包括:访问数据传输网络、访问数字照相机、访问麦克风等。由给定应用130

请求的许可可以在文件中描述,该文件可以是应用130的应用包的一部分。

[0051] 在一方面,可以使用数据结构领域中公知的任何方法将由应用130的可执行代码使用的类和方法的信息形式化为应用130的属性。例如,应用的属性可以被以具有多个级别的树状结构(列表)的形式提供:

[0052] • 类1(Class 1)

[0053] o方法1

[0054] o方法2

[0055] • 类2(Class 2)

[0056] o方法3

[0057] 然后,收集的应用130的属性集可以由系统保护器120发送到安全服务器150,并且具体地发送到存在于安全服务器150上的分类服务160。

[0058] 分类服务160被设计为基于应用的属性集来分类应用,尤其是应用130。

[0059] 应用130的分类结果被呈现为应用130属于应用的某些类别的概率。例如,应用130的分类结果可以如下表示:80%恶意应用、85%不需要的应用、60%可信应用。

[0060] 为了获取上述分类结果,分类服务160使用先前训练的专家分类系统。在一方面,训练的分类系统可从以下构建:神经网络、决策树或梯、贝叶斯分类器和/或本领域已知的任何其他分类器系统。

[0061] 为了分类服务160的训练,在一方面,可以使用标记应用集,其与应用类别的从属关系被认为是已知的。在另一方面,训练也可以使用在安全服务器150上运行的附加系统保护器。需要注意的是:标记应用集的形成可以由信息技术领域的专家或由本领域已知的任何专家系统来执行。

[0062] 一旦由分类服务160完成分类,应用130的分类结果就由安全服务器150发送到移动计算设备110中的系统保护器120。

[0063] 在一方面,系统保护器120使用从分类服务160接收的应用130的分类结果来确定应用130的类别。在一方面,使用启发式规则来执行应用的类别的确定,启发式规则被存储在反病毒记录的数据库140中。在一方面,启发式规则要求基于应用130的分类结果来确定应用130的类别。

[0064] 启发式规则的示例可以如下表示:

[0065] “如果对于应用,属于的概率:

[0066] • 属于恶意应用的类别的概率大于30%;以及

[0067] • 属于可信应用的类别的概率小于20%;

[0068] 然后,应用被分类为恶意的”。

[0069] 在又另一方面,系统保护器120使用启发式规则,例如,存储在数据库140中的规则,这些规则不仅被应用于应用130的分类结果,而且被应用于应用130的附加属性集,这些附加属性可以由系统保护器120收集。

[0070] 在一方面,应用130的附加属性集(也简称为“附加属性”)是与用于应用130的分类的属性完全相同的属性。在另一方面,该附加属性集是与分类服务160所使用的属性不同的属性。

[0071] 附加属性集的示例可以是:

[0072] • 来自应用包的可执行文件的大小；

[0073] • 来自应用包的可执行文件的指令序列。

[0074] 在一方面，系统保护器120使用以下中的至少一个来确定应用130的类别：从数据库140获取的启发式规则、分类结果和附加属性集。

[0075] 这种启发式规则的示例可以是："如果应用属于恶意应用类别的概率大于30％，同时应用包具有大小为100kB的可执行文件，并且该可执行文件包括包含三个异或(XOR)操作的序列的代码，则该应用被分类为恶意的"。

[0076] 在一方面，除了上述方法之外，系统保护器120还能够向安全服务器150发出获取应用130的分类结果的请求，而不向服务器150发送应用130的属性集。在这种情况下，该请求被寻址到服务器150，并且具体地被寻址到在安全服务器150内运行(操作)的信誉服务170，其中该请求本身可以包含应用130的标识符，诸如其名称、应用包的校验和、或唯一地表征应用130的任何其他标识符等。

[0077] 当请求被发送到信誉服务170时，为了向系统保护器120提供应用130的分类结果而不提供应用130的属性集，信誉服务170使用存储的先前分类的应用的分类结果。通过由分类服务160使用由系统保护器120先前收集的先前分类的应用的属性集执行分类，来获取先前分类的应用的分类结果。

[0078] 在一方面，信誉服务170利用判决数据库180来存储这些数据。此外，各种应用的分类结果可以结合它们各自的用于获取分类结果的情况存储在判决数据库180中。例如，所述情况可以包括：

[0079] • 分类的时间戳；

[0080] • 被分类的应用的标识符；

[0081] • 使用的专家分类系统的类型；

[0082] • 使用的专家分类系统的错误估计；

[0083] • 用于收集由系统保护器使用的属性的机制的版本(1.0、1.1、2.0、3.0等)；以及

[0084] • 专家分类系统的机制的版本(1.0、1.1、2.0、3.0等)。

[0085] 先前获取的应用分类结果包括由分类服务160应来自运行在移动计算设备110上的系统保护器120的请求而执行的分类结果，以及应那些来自可能位于一个或更多个用户的计算设备上或安全服务器150上的类似保护系统的请求而执行的分类结果。这种先前分类的应用可以是系统保护器(存在于设备110和服务器150上)已经收集了属性集并且将收集的属性发送到分类服务160以便获取分类结果的应用。

[0086] 在一方面，这样的应用是来自上述标记应用集的应用。

[0087] 无论何时分类服务160生成某些应用(包括应用130)的分类结果，服务160可以向信誉服务170发送分类结果。在一方面，用于获取分类结果的情况也被发送到信誉服务170。继而，信誉服务170可以将接收到的信息存储在判决数据库180中。

[0088] 由信誉服务170提供的应用130的分类的结果，包括获取的分类结果的情况，被发送到系统保护器120以便确定分类结果的相关性(作为信息的属性的相关性)。

[0089] 在一方面，分类结果的相关性由存在于移动计算设备110上的系统保护器120确定。在另一方面，分类结果的相关性由存在于安全服务器150上的另一系统保护器来确定。因此，应用130的分类结果的相关性的确定可以在移动计算设备110上执行或在安全服务器

150上执行。

[0090] 为了确定应用130的分类的相关性,系统保护器120(如同任何其他保护系统)使用用于确定相关性的规则(相关性规则)。这些相关性规则可以被存储在反病毒记录的数据库140中,该数据库可以包含要应用于应用130的分类结果的要求。

[0091] 如果应用130的分类结果满足相关性规则的要求,则分类结果被系统保护器120分类为相关的。

[0092] 相关性规则的要求的示例可以如下:

[0093] • 分类的时间戳和当前时间戳之间的差在允许值的指定范围内;

[0094] • 使用的专家分类系统的错误估计不大于0.5%;

[0095] • 使用的获取分类结果的专家分类系统的机制版本不低于在分类服务中使用的专家分类系统的机制版本;以及

[0096] • 由系统保护器为获取分类结果所使用的收集属性的机制的版本不低于在分类服务中使用的收集属性的机制的版本。

[0097] 在一方面,为了将应用130的分类结果指定为相关的,分类结果必须满足至少一个相关性规则。

[0098] 在另一方面,为了将应用130的分类结果指定为相关的,分类结果必须满足存储在反病毒记录的数据库140中的所有相关性规则。

[0099] 如果系统保护器120没有把应用130的分类结果分类(或指定)为相关的,在一方面,系统保护器120按照上述可能性收集用于发送到分类服务160的应用130的属性集,并且随后获取应用130的分类结果(例如,在同步模式下)。这样获取的分类结果被系统保护器120认为是相关的(即,没有借助于相关性规则的进一步相关性校对)。

[0100] 然后,在一方面,系统保护器120可使用应用130的分类的相关性结果来确定应用130的类别。

[0101] 在一方面,上述方案与诸如存储在反病毒记录的数据库140中的规则等启发式规则一起使用。

[0102] 在又另一方面,简化的方法可以用于确定应用130的类别。例如,应用130的类别可以被定义为基于分类结果具有应用130属于它的最高概率的类别。

[0103] 在一方面,被系统保护器120分类为恶意的应用可以被系统保护器120移除,或者可以被置于隔离区。在一方面,被分类为不需要的应用也可以被系统保护器120移除。在另一方面,当应用被分类为不需要的时,系统保护器120可以简单地向移动计算设备110的用户通知不需要的应用在设备110上的存在,和/或向用户建议一个或更多个选项,例如是否移除不需要的应用的选项。

[0104] 在另一方面,系统保护器120可以撤销授予应用130的许可,以便保护移动计算设备110免受可以由应用130执行的动作的影响。

[0105] 应当注意,反病毒记录的数据库140和判决数据库180的内容可以由信息技术领域中的专家修改。此外,数据库140的内容也可以通过从安全服务器150接收的命令来远程修改。

[0106] 图2示出了根据本公开的教导的用于使用分类服务来对应用进行分类的方法200。方法200可以被在包括任何数量的设备的计算系统上实施,例如,包括移动计算设备110和

安全服务器150的计算系统100。

[0107] 步骤201,方法200通过运行在用户的移动计算设备110上的系统保护器120收集存在于设备110上的应用130的属性集。

[0108] 步骤202,方法200通过系统保护器120将所收集的属性集发送到安全服务器150,更准确地说,发送到存在于安全服务器150上的分类服务160。

[0109] 步骤203,方法200通过存在于安全服务器150上的分类服务160,基于从系统保护器120接收的属性集来对应用130进行分类。在一方面,分类的结果包括:应用130属于应用的一个或更多个相应类别的一个或更多个概率。在一方面,应用的类别包括:可信的、恶意的和不需要的。分类结果由分类服务160发送到存在于设备110上的系统保护器120。

[0110] 步骤204,方法200通过系统保护器120基于从分类服务160接收的分类结果确定应用的类别。

[0111] 在一方面,进一步基于应用130的属性确定应用的类别。在一方面,用于确定应用的类别的应用的属性包括以下中的至少一个:从应用收集的属性集,以及附加属性集,附加属性是不用于应用130的分类的属性。

[0112] 当如上所述的方法200用于确定应用130的类别时,使用应用的属性(例如使用本地存储在移动计算设备110上的启发式规则)来细化应用130的分类结果。因此,方法200有助于描述在对应用130进行分类时的I型和II型错误(假阳性和遗漏)。

[0113] 此外,如果错误发生时,本地存储的启发式规则通过更新启发式规则来使错误能够快速修正。因此,与需要对分类算法进行复杂的再训练的专家分类系统不同,本地存储的启发式规则中的错误可以在短时间内并且在没有复杂系统的情况下被修正。换句话说,当发现错误时,可以在本地执行启发式规则的更新,从而以更灵敏的方式改进应用的分类。

[0114] 应当注意,为了达到上述的改进,系统保护器120可以向安全服务器150发送关于应用130的任何给定信息和应用被通过系统保护器120分配的类别,这对于通过在本技术领域内通常已知的任何方法检测第一和第二类错误是必要的。在检测到第一或第二类错误之后,安全服务器150(例如借助于运行在服务器150上的系统保护器120)能够向系统保护器120提供对启发式规则的改变。对启发式规则的改变可以包括对特定规则的改变或对用于发出删除规则的指令的改变。对启发式规则进行改变,以防止在使用来自数据库140的修改后的本地启发式规则时,应用130的类别的错误确定。此外,当在启发式规则的修正之后使用专家分类系统来确定应用的类别时,本公开的方法实现了第一和第二类错误的更快的修正。

[0115] 图3示出了根据本公开的教导的用于基于相关性规则对计算设备上的应用进行分类的方法300。方法300可以在包括任何数量的设备的计算系统上实施,例如,包括移动计算设备110和安全服务器150的计算系统100。

[0116] 步骤301,方法300由系统保护器120发送请求以获取应用130的分类结果。在一方面,该请求被发送到安全服务器150,尤其是发送到存在于安全服务器150上的信誉服务170。信誉服务170存储分类应用的分类的结果,即,由分类服务160执行的任何数量的应用的先前分类的结果。先前的分类基于由系统保护器120收集的先前分类的应用集合。

[0117] 步骤302,方法300由系统保护器120获取应用130的分类结果。例如,系统保护器120从信誉服务170接收分类的结果。

[0118] 步骤303,方法300由系统保护器120基于相关性规则确定所获取的应用130的分类结果是否为相关的。当系统保护器120认为应用的分类结果是相关的时,方法300前往至步骤320。在一方面,当系统保护器120没有认为应用130的分类结果是相关的时,系统保护器120前往至步骤310,在另一方面,系统保护器120简单地终止方法300。

[0119] 步骤310,方法300执行关于方法200结合图2描述的步骤。因此,系统保护器120收集并向安全服务器150发送应用130的属性集。然后,系统保护器120从安全服务器150获取来自应用130的分类的更新结果。

[0120] 步骤320,在一方面,通过系统保护器130,方法300可以基于相关性规则将应用130的分类的更新结果指定为相关的。满足相关性规则的应用130的分类结果被系统保护器120认为是相关的。

[0121] 步骤330,通过系统保护器120,方法300基于被指定为关联的分类结果确定应用130的类别。

[0122] 在一方面,分类的更新包括响应于向安全服务器发送属性集而从安全服务器接收更新的分类。

[0123] 在一方面,应用的属性集包括以下中的至少一个:在应用的应用包中的多个文件、在应用包中的多个可执行文件、多个被请求的许可以及许可的类型、应用包中的可执行文件中的多个类、以及应用包中的可执行文件中的多个方法。

[0124] 在一方面,应用的分类的结果被呈现为应用属于相应的应用的一个或更多个类别的一个或更多个概率。

[0125] 在一方面,应用的一个或更多个类别包括以下中的至少一个:恶意应用的类别、不需要的应用的类别和可信应用的类别。

[0126] 在一方面,使用至少部分基于分类结果的启发式规则来确定应用的类别。

[0127] 在一方面,该方法还包括:当应用被分类为恶意时,从计算设备移除该应用或隔离该应用,并且当应用被分类为不需要时,执行以下中的至少一个:从计算设备移除该应用、向计算设备的用户通知该不需要的应用在计算设备上的存在、向计算设备的用户提供用于选择保留或移除不需要的应用的选项,以及撤销先前授予该应用的许可。

[0128] 上述方法在减少分类服务160的负担的方面是有利的,并且因此也减少了安全服务器150的负担,从而呈现了分类结果的改进。此外,确定由信誉服务170提供的先前执行的分类的结果的相关性有利地减少了在应用130的分类期间的第一和第二类错误,并因此也在确定应用130的类别时,由于“过时”,错误的和不相关的分类将不被用于确定应用130的类别。

[0129] 图4示出了根据示例性方面的可以在其上实施的对计算设备上的应用进行分类的系统和方法的各方面的计算机系统20的框图。应当注意,计算机系统20可对应于例如如前所述的虚拟安全设备102。计算机系统20可以是多个计算设备的形式,或者是单个计算设备的形式,例如,台式计算机、笔记本计算机、便携计算机、移动计算设备、智能电话、平板计算机、服务器、大型机、嵌入式设备和其他形式的计算设备。

[0130] 如图所示,计算机系统20包括中央处理单元(central processing unit,CPU)21、系统存储器22和连接各种系统组件的系统总线23,系统组件包括与中央处理单元21相关联的存储器。系统总线23可以包括总线存储器或总线存储控制器、外围总线和能够与任何其

它总线架构交互的局部总线。总线的示例可以包括外部控制器接口(PCI)、工业标准结构总线(ISA)、高速串行计算机总线(PCI-Express)、HT总线(HyperTransport™)、无限带宽(InfiniBand™)、串行ATA(SerialATA)、I2C(I^2C)和其它合适的连接器产品。中央处理单元21(也称为处理器)包括具有单个或多个内核的单个或多个处理器组。处理器21可以执行实施本公开的技术的一个或更多个计算机可执行代码。系统存储器22可以是用于存储本文所使用的数据和/或可由处理器21执行的计算机程序的任何存储器。系统存储器22可以包括诸如随机存取存储器(RAM)25等易失性存储器和诸如只读存储器(ROM)24的非易失性存储器、闪存,或其任意组合。基本输入/输出系统(BIOS)26可以存储用于在计算机系统20的元件之间传输信息的基本程序,诸如在使用ROM 24加载操作系统时的那些程序。

[0131] 计算机系统20可以包括一个或更多个诸如一个或更多个可移动存储设备27、一个或更多个固定存储设备28或其组合等存储设备。一个或更多个可移动存储设备27和固定存储设备28经由存储接口32连接到系统总线23。在一方面,存储设备和对应的计算机可读存储介质是用于存储计算机系统20的计算机指令、数据结构、程序模块和其他数据的功率独立模块。系统存储器22、可移动存储设备27和固定存储设备28可以使用各种计算机可读存储介质。计算机可读存储介质的示例包括机械存储器,诸如高速缓存、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、零电容器RAM、双晶体管RAM、增强动态随机存取存储器(eDRAM)、动态存储器(EDO RAM)、双倍数据速率存储器(DDR RAM)、电可擦除只读存储器(EEPROM)、纳米管随机存取存储器(NRAM)、电阻式随机存储器(RRAM)、半导体(SONOS)、相变存储器(PRAM);闪存或诸如在固态驱动器(SSD)或闪存驱动器中的其他存储器技术;诸如在硬盘驱动器或软盘中的磁盒、磁带和磁盘存储;诸如光盘(CD-ROM)或数字多功能盘(DVD)中的光存储器;以及可用于存储所需数据并可由计算机系统20存取的任何其它介质。

[0132] 计算机系统20的系统存储器22、可移动存储设备27和固定存储设备28可用于存储操作系统35、附加程序应用37、其它程序模块38和程序数据39。计算机系统20可以包括外围接口46,该外围接口46用于经由一个或更多个I/O端口(诸如串行端口、并行端口、通用串行总线(USB)或其他外围接口)从输入设备40(诸如键盘、鼠标、触控笔、游戏控制器、语音输入设备、触摸输入设备或其他诸如打印机或扫描仪等外围设备等)传送数据。诸如一个或更多个监视器、投影仪或集成显示器等显示设备47也可通过诸如视频适配器等输出接口48连接到系统总线23。除了显示设备47之外,计算机系统20可以配备有其他外围输出设备(未示出),诸如扬声器和其他视听设备等。

[0133] 计算机系统20可以使用到一个或更多个远程计算机49的网络连接在网络环境中操作。一个远程计算机(或多个远程计算机)49可以是本地计算机工作站或服务器,该本地计算机工作站或服务器包括描述计算机系统20的性质中的大部分或全部上述元件。计算机网络中还可以存在其他设备,诸如但不限于路由器、网络站、对等设备或其他网络节点等。计算机系统20可以包括一个或更多个网络接口51或网络适配器,该网络接口51或网络适配器用于经由一个或更多个网络与远程计算机49通信,该网络诸如局域计算机网络(LAN)50、广域计算机网络(WAN)、内联网和因特网等。网络接口51的示例可以包括以太网接口、帧中继接口、同步光纤网络接口(SONET接口)和无线接口。

[0134] 本公开的方面可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括具有计算机可读程序指令的计算机可读存储介质(或媒介),该计算机可读程序指令用于使

处理器执行本公开的各方面。

[0135] 计算机可读存储介质是能够保持和存储以指令或数据结构的形式的程序代码的有形设备,所述指令或数据结构被通过诸如计算系统20等计算设备的处理器存取。计算机可读存储介质可以是电子存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或其任何合适的组合。作为示例,这样的计算机可读存储介质包括随机存取存储器(RAM)、只读存储器(ROM)、电可擦除只读存储器(EEPROM)、便携式光盘只读存储器(CD-ROM)、数字多功能盘(DVD)、闪存、硬盘、便携式计算机磁盘、记忆棒、软盘、或者甚至是诸如打孔卡或在凹槽中具有指令记录的凸起结构等机械编码设备。如本文所使用的,计算机可读存储介质不应被解释为本身是短暂的信号,诸如无线电波或其他自由传播的电磁波、通过波导或传输介质传播的电磁波、或通过导线传输的电信号等。

[0136] 本文描述的计算机可读程序指令从计算机可读存储介质下载到相应的计算设备,或者经由网络(例如因特网、局域网、广域网和/或无线网络)下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光传输光纤、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算设备中的网络接口从网络接收计算机可读程序指令,并转发计算机可读程序指令以存储在相应的计算设备内的计算机可读存储介质中。

[0137] 用于执行本公开的操作的计算机可读程序指令可以是汇编指令、指令集架构(ISA)指令、机械指令、机械相关指令、微代码、固件指令、状态设置数据、或者以一种或更多种编程语言的任意组合编写的源代码或对象代码,该编程语言包括面向对象的编程语言和常规的程序编程语言。计算机可读程序指令可以完全在用户的计算机上执行,部分在用户的计算机上执行,作为独立的软件包执行,部分在用户的计算机上并且部分在远程计算机上执行,或者完全在远程计算机或服务器上执行。在后一种情境下,远程计算机可以通过任何类型的网络(包括LAN或WAN)连接到用户的计算机,或者可以连接到外部计算机(例如,通过因特网)。在一些方面,包括例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA)的电子电路可以通过利用计算机可读程序指令的状态信息来执行计算机可读程序指令以使电子电路个性化,以便执行本公开的方面。

[0138] 在各个方面,可以按照模块来解决本公开的系统和方法。这里使用的术语“模块”指的是使用诸如通过专用集成电路(application specific integrated circuit,ASIC)或现场可编程门阵列(FPGA)等硬件实施的真实世界的设备、组件或组件的布置,例如,或者作为硬件和软件的组合,诸如通过微处理器系统和指令集来实现模块的功能,其(在被执行时)将微处理器系统转换为专用设备。模块也可以作为两者的组合被实施,其中某些功能单独由硬件实现,而其它功能由硬件和软件的组合实现。在某些实施例中,模块的至少一部分,并且在某些情况下,模块的全部可以被在计算机系统(诸如以上在图4中更详细的描述的计算机系统)的处理器上执行。因此,每个模块可以以各种合适的配置来实现,并且不应限于本文的任何特定的示例性实施。

[0139] 为了清楚起见,本文中未公开各方面的所有常规特征。应当理解,在本公开的任何实际实施方式的发展中,必须做出许多实施方式特定的决定以便实现开发者的特定目标,并且这些特定目标将针对不同的实施方式和不同的开发者而变化。应当理解,这种开发工作可能是复杂的且耗时的,但是对于受益于本公开的本领域普通技术人员来说,这不过是常规的工程任务。

[0140] 此外,应当理解,本文所使用的措辞或术语是出于描述而非限制的目的,使得本说明书的术语或措辞由本领域技术人员根据本文呈现的教导和指导,结合(一个或更多个)相关领域的技术人员的知识来理解。此外,除非这样明确地陈述否则不意在将说明书或权利要求书中的任何术语归结为不常见的或特殊的含义。

[0141] 本文公开的各个方面包含了本文通过示例的方式提及的已知模块的当前和未来所知的等同物。此外,虽然已经示出和描述了各方面和应用,但是对于受益于本公开的本领域技术人员而言将显而易见的是,在不背离本文所公开的发明思想的情况下,比上述更多的修改是可能的。

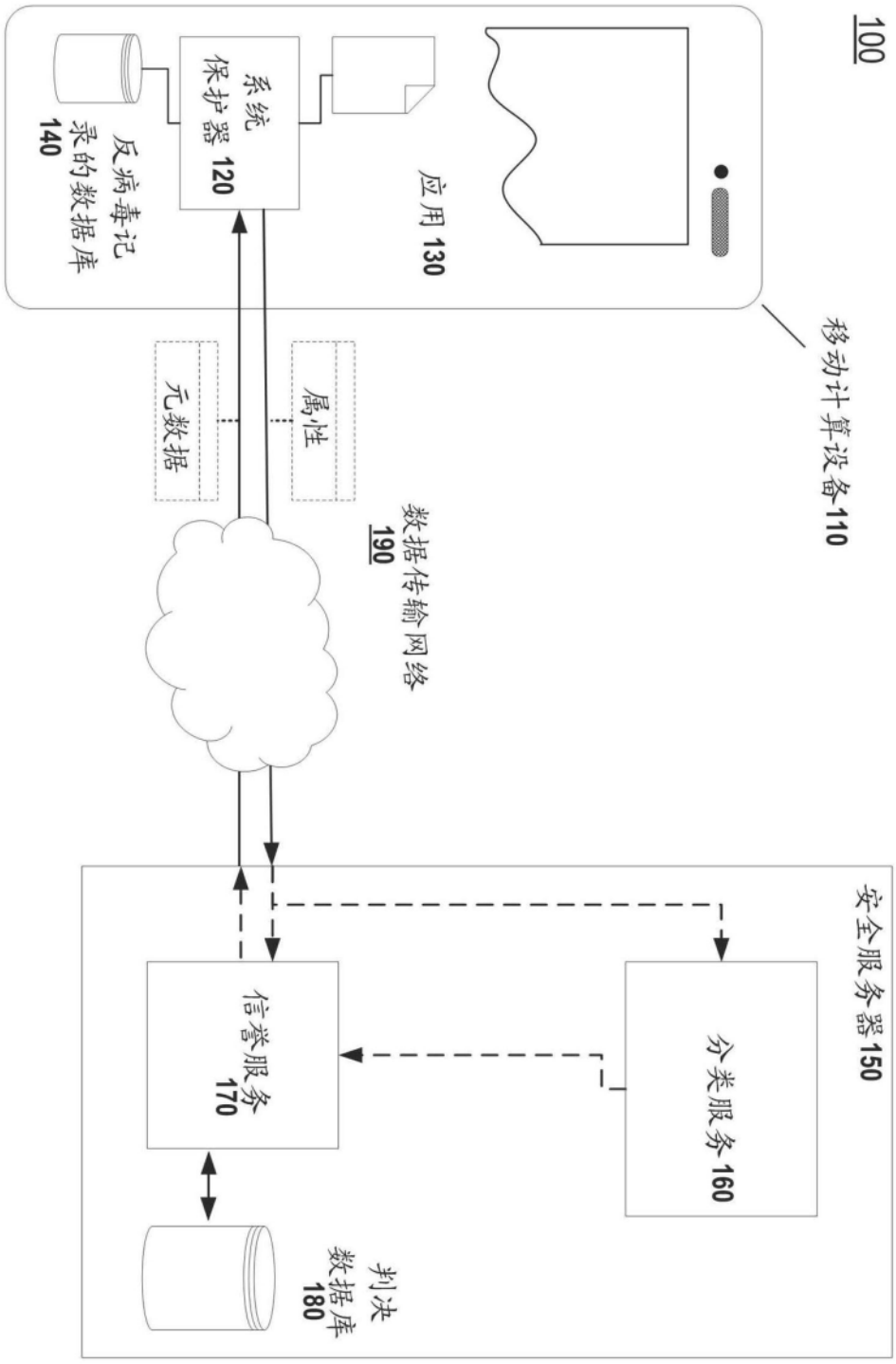

100
移动计算设备110
应用130
系统保护器120
反病毒记录的数据库140
元数据
属性
数据传输网络190
安全服务器150
分类服务160
信誉服务170
判决数据库180

图1

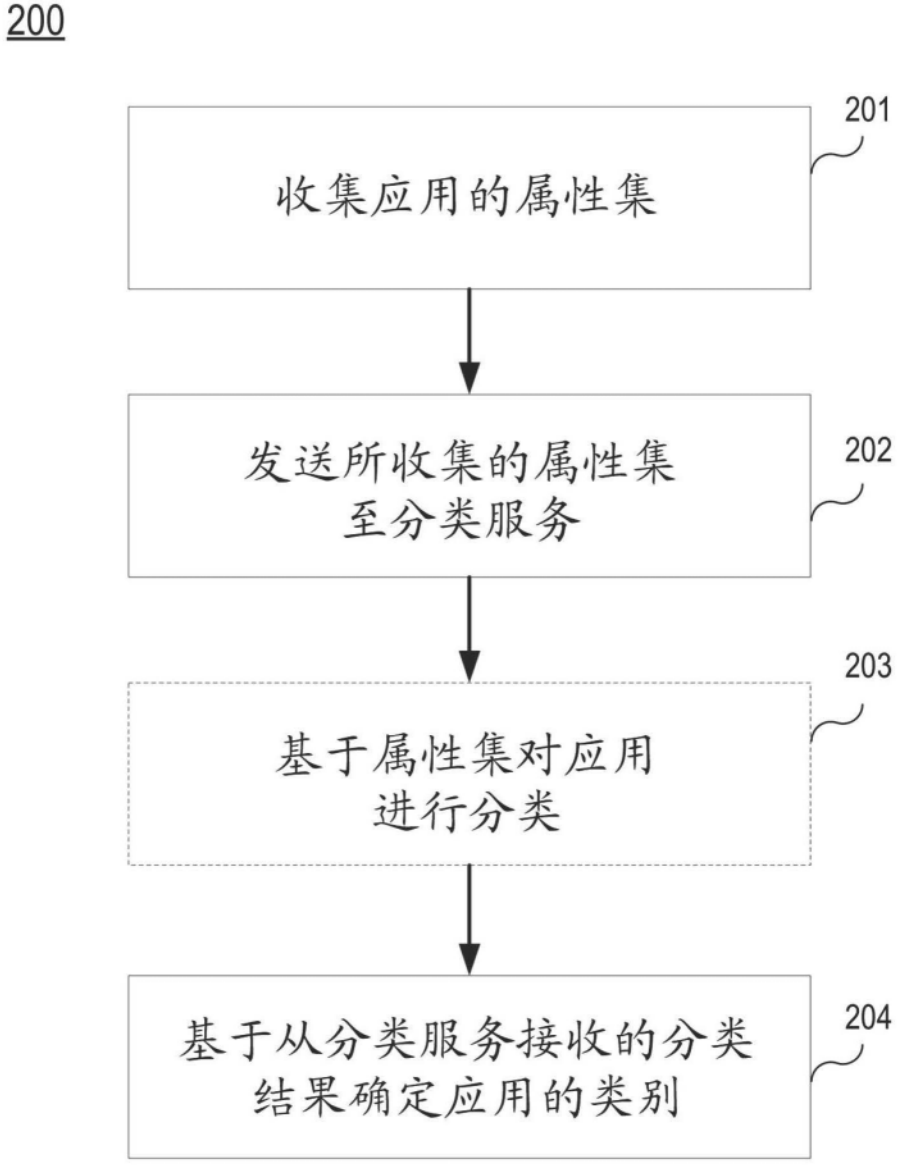

200
201
收集应用的属性集
202
发送所收集的属性集
至分类服务
203
基于属性集对应用
进行分类
204
基于从分类服务接收的分类
结果确定应用的类别

图2

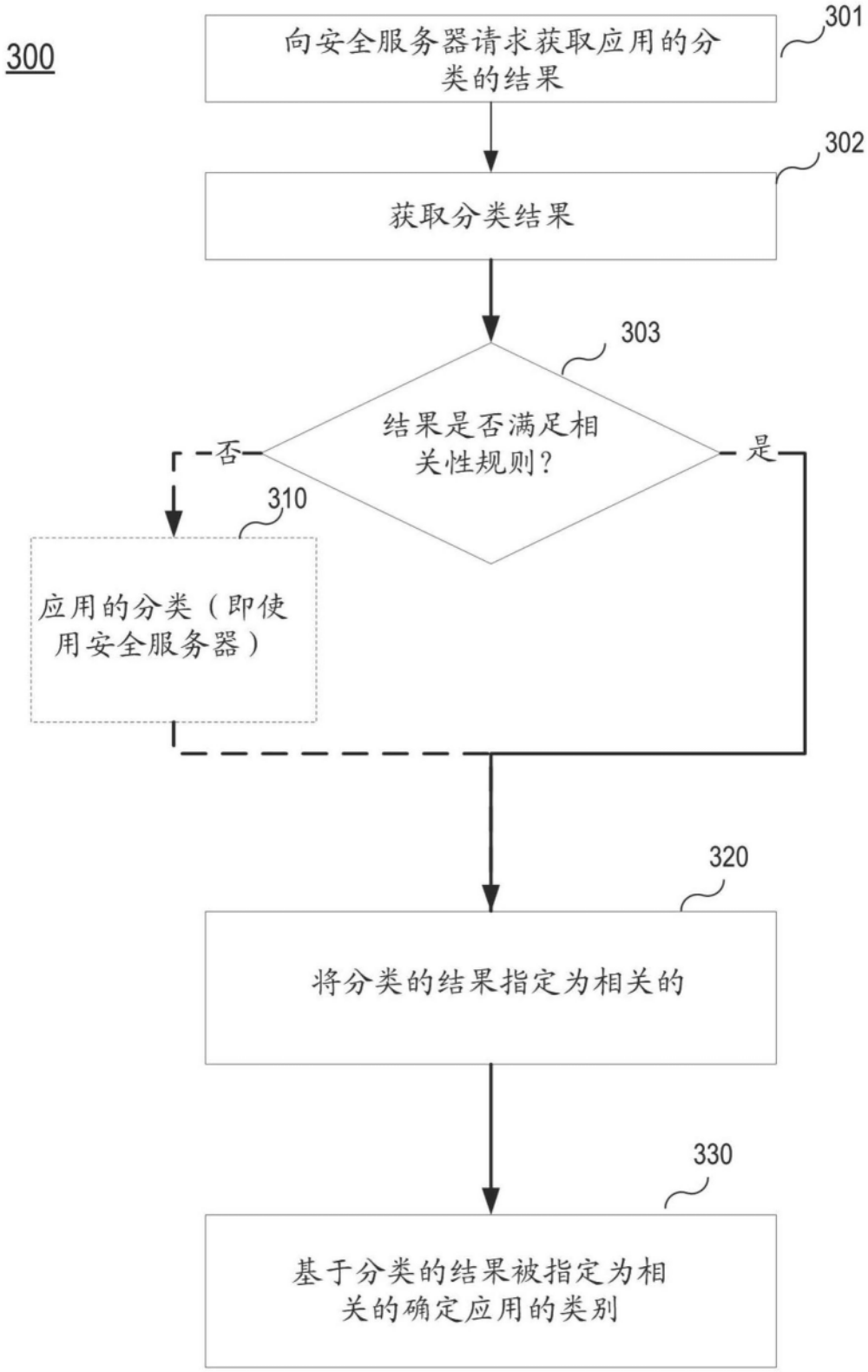

300
向安全服务器请求获取应用的分类的结果
301
获取分类结果
302
结果是否满足相关性规则？
303
否
是
应用的分类（即使用安全服务器）
310
将分类的结果指定为相关的
320
基于分类的结果被指定为相关的确定应用的类别
330

图3

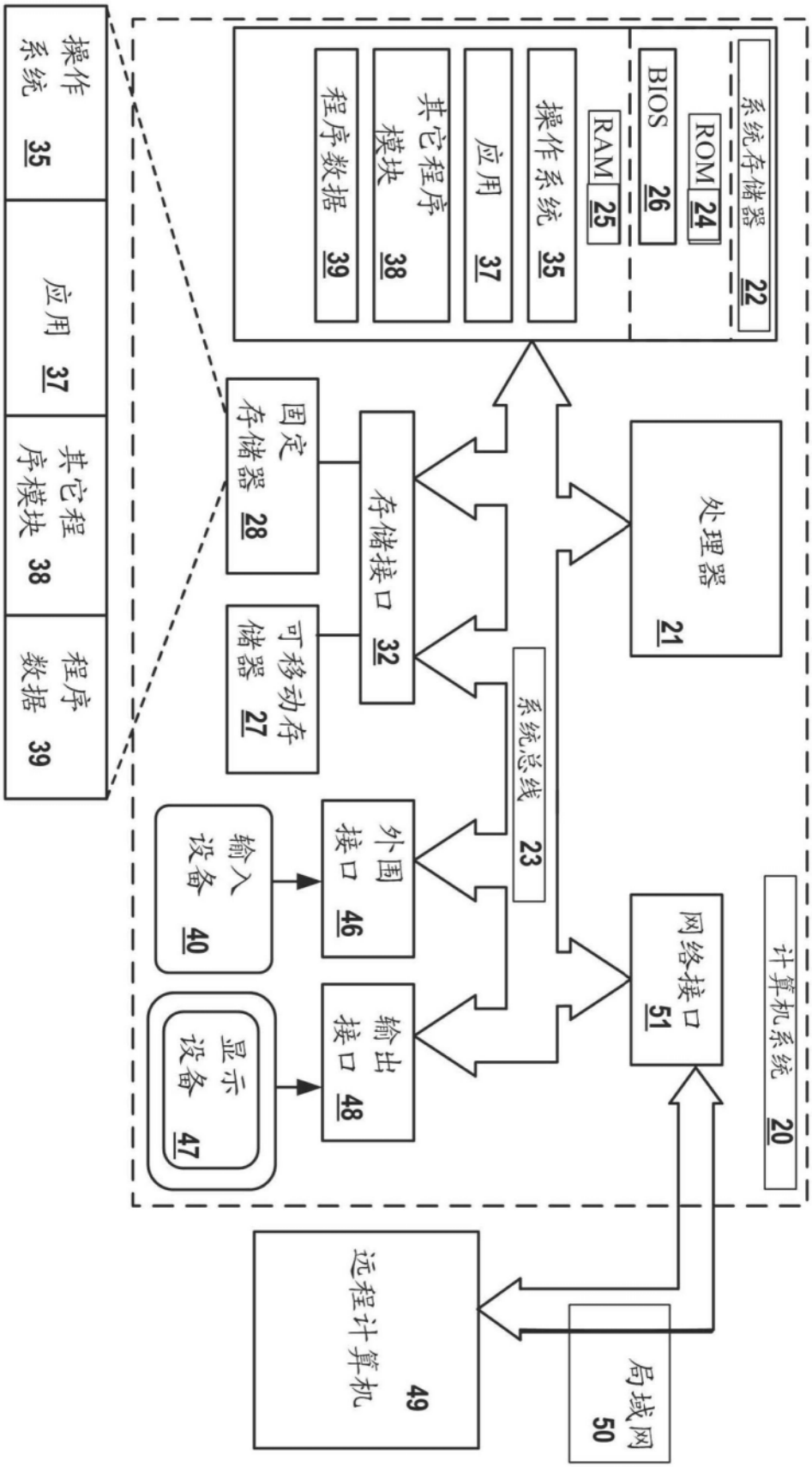

计算机系统 20
系统存储器 22
ROM 24
BIOS 26
RAM 25
操作系统 35
应用 37
其它程序模块 38
程序数据 39
处理器 21
系统总线 23
存储接口 32
固定存储器 28
可移动存储器 27
外围接口 46
输入设备 40
输出接口 48
显示设备 47
网络接口 51
局域网 50
远程计算机 49
操作系统 35
应用 37
其它程序模块 38
程序数据 39

图4