(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 101884047 A
(43)申请公布日 2010.11.10

(21)申请号 200880119072.0
(22)申请日 2008.10.03
(30)优先权数据
11/868,321 2007.10.05 US
12/041,309 2008.03.03 US
(85)PCT申请进入国家阶段日
2010.06.03
(86)PCT申请的申请数据
PCT/US2008/078807 2008.10.03
(87)PCT申请的公布数据
WO2009/046340 EN 2009.04.09
(71)申请人 谷歌公司
地址 美国加利福尼亚州
(72)发明人 周云凯 尼尔斯·普罗沃斯
小克莱顿·W·巴沃尔
埃里克·L·戴维斯
马克·帕拉图奇
卡玛尔·P·尼加姆
克里斯托弗·K·蒙森
帕纳约蒂斯·马弗洛马蒂斯
雷切尔·那卡乌奇
(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 张焕生 安翔
(51)Int.Cl.
G06F 21/00(2006.01)
G06Q 50/00(2006.01)

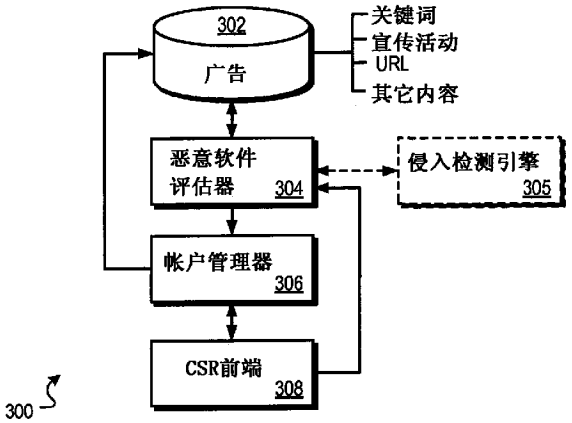
权利要求书 5 页 说明书 15 页 附图 9 页

(54)发明名称

侵入软件管理

(57)摘要

识别与赞助内容相关联的着陆页的侵入特征。基于所识别的侵入特征生成着陆页的特征分值，以及如果着陆页的特征分值超过特征阈值，则将着陆页分类为候选着陆页。可以暂停与候选着陆页相关联的赞助者帐户，或可以暂停与候选着陆页相关联的赞助内容。



CN 101884047 A

1. 一种方法，包括：

识别与赞助内容相关联的着陆页；

识别所述着陆页的侵入特征；

基于所识别的侵入特征生成所述着陆页的特征分值；

确定所述着陆页的所述特征分值是否超过特征阈值；

如果所述着陆页的所述特征分值超过所述特征阈值，则将所述着陆页分类为候选着陆页；

将所述候选着陆页提交给侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的侵入分值；以及

如果所述侵入分值超过侵入阈值，则排除对与所述候选着陆页相关联的所述赞助内容的派发。

2. 如权利要求 1 所述的方法，其中所述赞助内容是广告。

3. 如权利要求 2 所述的方法，进一步包括：

接收对所述赞助者帐户的上诉请求，并且响应于接收所述上诉请求：

将所述候选着陆页重新提交给所述侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的另一个侵入分值；

如果所述另一个侵入分值超过所述侵入阈值，则排除对与所述候选着陆页相关联的所述广告的派发；以及

如果所述另一个侵入分值没有超过所述侵入阈值，则允许对与所述候选着陆页相关联的所述广告的派发。

4. 如权利要求 3 所述的方法，进一步包括：

识别与所述广告相关联的赞助者帐户，所述赞助者帐户包括额外广告；以及

如果所述候选着陆页的所述侵入分值超过所述侵入阈值，则排除对与所述赞助者帐户相关联的所述额外广告的派发。

5. 如权利要求 4 所述的方法，进一步包括：

接收对所述赞助者帐户的上诉请求，并且响应于接收所述上诉请求：

将所述候选着陆页和与所述额外广告相关联的额外着陆页提交给所述侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的另一个侵入分值和所述额外着陆页的额外侵入分值；以及

确定所述着陆页的所述侵入分值是否超过所述侵入阈值。

6. 如权利要求 5 所述的方法，进一步包括：排除对与具有超过所述侵入阈值的侵入分值的着陆页相关联的广告的派发。

7. 如权利要求 5 所述的方法，进一步包括：如果着陆页中的任何着陆页的侵入分值超过所述侵入阈值，则排除对与所述赞助者帐户相关联的广告的派发。

8. 如权利要求 2 所述的方法，其中所述侵入特征包括一个或多个 iFrame 特征、一个或多个 URL 特征、和 / 或一个或多个脚本特征，并且其中每一个侵入特征具有关联的特征权重，以及所述基于所识别的侵入特征生成所述着陆页的侵入分值包括对所述特征权重求和。

9. 如权利要求 2 所述的方法，进一步包括：

识别与所述广告相关联的赞助者帐户，所述赞助者帐户包括额外广告；

识别与所述额外广告相关联的额外着陆页；

基于所识别的侵入特征生成所述额外着陆页的特征分值；

确定所述着陆页的所述特征分值是否超过特征阈值；

如果所述额外着陆页的所述特征分值超过所述特征阈值，则将每一个额外着陆页分类为额外候选着陆页；

累积所述候选着陆页和所述额外候选着陆页的所述特征分值以获取帐户分值；

基于所述帐户分值将风险类别分配给所述赞助者帐户；以及

基于所述风险类别来启动用于所述赞助者帐户的多个帐户补救过程中的一个。

10. 一种方法，包括：

将与广告相关联的着陆页分割成训练着陆页和测试着陆页；

针对所述训练着陆页的侵入特征迭代训练分类模型；

针对所述测试着陆页的所述侵入特征迭代测试所述分类模型直到出现测试停止事件；以及

响应于所述停止事件将特征权重和侵入特征的关联存储在所述分类模型中，所述特征权重和侵入特征的关联取自所述迭代训练和测试。

11. 如权利要求 10 所述的方法，其中所述分类模型包括基于线性回归的模型。

12. 如权利要求 10 所述的方法，进一步包括：

识别与广告相关联的着陆页；

识别所述着陆页的侵入特征；

基于所识别的侵入特征和特征权重生成所述着陆页的特征分值；

确定所述着陆页的所述特征分值是否超过特征阈值；以及

如果所述着陆页的所述特征分值超过所述特征阈值，则将所述着陆页分类为候选着陆页。

13. 如权利要求 12 所述的方法，进一步包括：

将所述候选着陆页提交给侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的侵入分值；以及

如果所述侵入分值超过侵入阈值，则排除对与所述候选着陆页相关联的所述广告的派发。

14. 如权利要求 13 所述的方法，进一步包括：

接收对所述赞助者帐户的上诉请求，并且响应于接收所述上诉请求：

将所述候选着陆页重新提交给所述侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的另一个侵入分值；

如果所述另一个侵入分值超过所述侵入阈值，则排除对与所述候选着陆页相关联的所述广告的派发；以及

如果所述另一个侵入分值没有超过所述侵入阈值，则允许对与所述候选着陆页相关联的所述广告的派发。

15. 如权利要求 13 所述的方法，进一步包括：

识别与所述广告相关联的赞助者帐户，所述赞助者帐户包括额外广告；以及

如果所述候选着陆页的所述侵入分值超过所述侵入阈值，则排除对与所述赞助者帐户相关联的所述额外广告的派发。

16. 如权利要求 10 所述的方法，其中所述侵入特征包括一个或多个 iFrame 特征、一个或多个 URL 特征、和 / 或一个或多个脚本特征。

17. 一种系统，包括：

评分引擎，所述评分引擎包括存储在计算机可读介质中并且由处理系统可执行的软件指令，以及在这样的执行时促使所述处理系统：

识别与赞助内容相关联的着陆页；

识别所述着陆页的侵入特征；

基于所识别的侵入特征生成所述着陆页的特征分值；

确定所述着陆页的所述特征分值是否超过特征阈值；

如果所述着陆页的所述特征分值超过所述特征阈值，则将所述着陆页分类为候选着陆页；以及

恶意软件评估引擎，所述恶意软件评估引擎包括存储在计算机可读介质中并且由处理系统可执行的软件指令，以及在这样的执行时促使所述处理系统：

将所述候选着陆页提交给侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的侵入分值；以及

如果所述侵入分值超过侵入阈值，则排除对与所述候选着陆页相关联的所述赞助内容的派发。

18. 如权利要求 17 所述的系统，其中所述赞助内容是广告。

19. 如权利要求 18 所述的系统，进一步包括：

上诉处理评估引擎，所述上诉处理评估引擎包括存储在计算机可读介质中并且由处理系统可执行的软件指令，以及在这样的执行时促使所述处理系统：

接收对与所述着陆页相关联的赞助者帐户的上诉请求，并且响应于所述上诉请求促使所述恶意软件评估引擎：

将所述候选着陆页重新提交给所述侵入检测引擎；

从所述侵入检测引擎接收所述候选着陆页的另一个侵入分值；

如果所述另一个侵入分值超过所述侵入阈值，则排除对与所述候选着陆页相关联的所述广告的派发；以及

如果所述另一个侵入分值没有超过所述侵入阈值，则允许对与所述候选着陆页相关联的所述广告的派发。

20. 如权利要求 18 所述的系统，其中所述恶意软件评估引擎包括存储在计算机可读介质中并且由处理系统可执行的进一步软件指令，以及在这样的执行时促使所述处理系统：

识别与所述广告相关联的赞助者帐户，所述赞助者帐户包括额外广告；以及

如果所述候选着陆页的所述侵入分值超过所述侵入阈值，则排除对与所述赞助者帐户相关联的所述额外广告的派发。

21. 如权利要求 20 所述的系统，进一步包括：

上诉处理评估引擎，所述上诉处理评估引擎包括存储在计算机可读介质中并且由处理

系统可执行的软件指令，以及在这样的执行时促使所述处理系统：

接收对与所述着陆页相关联的赞助者帐户的上诉请求，并且响应于所述上诉请求促使所述恶意软件评估引擎：

识别与所述赞助者帐户相关联的着陆页；

将所述着陆页提交给所述侵入检测引擎；

从所述侵入检测引擎接收所述着陆页的侵入分值；以及

如果所述侵入分值中的每一个没有超过所述侵入阈值，则允许对所述广告和所述额外广告的派发。

22. 如权利要求 21 所述的系统，进一步包括：

数据存储，所述数据存储存储与广告相关联的训练着陆页和与广告相关联的测试着陆页；以及

机器学习引擎，所述机器学习引擎包括存储在计算机可读介质中并且由处理系统可执行的软件指令，以及在这样的执行时促使所述处理系统：

针对所述训练着陆页的侵入特征迭代训练分类模型；

针对所述测试着陆页的所述侵入特征迭代测试所述分类模型直到出现测试停止事件；以及

响应于所述停止事件将特征权重和侵入特征的关联存储在所述分类模型中，所述特征权重和侵入特征的关联取自所述迭代训练和测试；

其中对所述着陆页的特征分值的生成基于存储在所述分类模型中的与所识别的侵入特征相关联的所述特征权重。

23. 一种系统，包括：

用于生成将特征权重与侵入特征相关联的分类模型的装置；

用于识别与广告相关联的候选着陆页的装置；

用于将所述候选着陆页提交给侵入检测引擎、用于接收所述候选着陆页的侵入分值以及用于排除对与所述候选着陆页相关联的赞助内容的派发的装置；以及

用于对排除对与被确定与侵入软件相关联的候选着陆页相关联的赞助内容的派发进行上诉的装置。

24. 一种方法，包括：

识别与赞助内容相关联的着陆页；

识别所述着陆页的侵入特征；

基于所述侵入特征的存在，评估所述着陆页以确定所述着陆页包括恶意软件的可能性是否超过第一阈值；

如果所述着陆页包括恶意软件的所述可能性超过所述第一阈值，则进一步评估所述着陆页，所述进一步评估包括针对所述着陆页执行至少一个额外的评估过程；以及

基于所述进一步评估，将所述着陆页分类为或包括恶意软件或不包括恶意软件的着陆页。

25. 如权利要求 24 所述的方法，进一步包括：

针对训练着陆页的侵入特征迭代训练分类模型；

针对测试着陆页的所述侵入特征迭代测试所述分类模型直到出现测试停止事件；以及

响应于所述停止事件将特征权重和侵入特征的关联存储在所述分类模型中，所述特征权重和侵入特征的关联取自所述迭代训练和测试并且用于促进对所述着陆页的所述评估。

26. 一种系统，包括：

数据存储，所述数据存储存储与广告相关联的训练着陆页和与广告相关联的测试着陆页；以及

机器学习引擎，所述机器学习引擎包括存储在计算机可读介质中并且由处理系统可执行的软件指令，以及在这样的执行时促使所述处理系统：

针对所述训练着陆页的侵入特征迭代训练分类模型；

针对所述测试着陆页的所述侵入特征迭代测试所述分类模型直到出现测试停止事件；以及

响应于所述停止事件将特征权重和侵入特征的关联存储在分类模型中，所述特征权重和侵入特征的关联取自所述迭代训练和测试。

27. 一种系统，包括：

用于将与广告相关联的着陆页分割成训练着陆页和测试着陆页的装置；

用于针对所述训练着陆页的侵入特征迭代训练分类模型和针对所述测试着陆页的所述侵入特征迭代测试所述分类模型直到出现测试停止事件，以及用于响应于所述停止事件将特征权重和侵入特征的关联存储在所述分类模型中的装置，所述特征权重和侵入特征的关联取自所述迭代训练和测试。

侵入软件管理

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2007 年 10 月 5 日提交的美国申请序列号 11/868,321 和于 2008 年 3 月 3 日提交的美国申请序列号 12/041,309 的优先权，在此以引用的方式合并上述申请所公开的内容。

技术领域

[0003] 本文档涉及侵入软件的管理。

背景技术

[0004] 交互式媒介（例如，因特网）具有对将例如广告（“ad”）的赞助内容定向到受众进行改进的巨大潜力。例如，一些网站提供基于由寻找信息的用户所输入的关键词的信息搜索功能。该用户查询可以是用户感兴趣的信息类型的指示符。通过将用户查询与由广告主所指定的关键词列表进行比较，可以向用户提供定向广告。

[0005] 在线广告的另一种形式是广告联合 (ad syndication)，其允许广告主通过将广告分发到其他伙伴来扩展他们的市场范围。例如，第三方在线发布者可以将广告主的文本或图像广告置放在具有期望内容的 web 资产上，以促使在线顾客到该广告主的网站。

[0006] 诸如包括数行文本、图像或视频剪辑的创意的广告包括指向着陆页 (landing page) 的链接。这些着陆页是在广告主网站或联合发布者网站上的页面，当用户在广告上点击时用户被指引到所述页面。然而，这些着陆页中的一些可能包括侵入软件，例如欺骗地、秘密地和 / 或自动地安装的软件、脚本或任何其它实体。侵入安装的这样的软件实体通常可以被表征为“恶意软件 (malware)”，单词“恶意的 (malicious)”和“软件 (software)”的混合词。然而，软件无需进行恶意行为才成为恶意软件；侵入安装的任何软件可以被认为是恶意软件，而不管软件进行的行为是否是恶意的。因此，除特洛伊木马、病毒以及浏览器漏洞外，诸如监视软件的其它软件也可以被认为是恶意软件。恶意软件可以有意或无意地存在于着陆页中。例如，广告主的站点能够被黑客攻击并且恶意软件直接插入到着陆页上；恶意广告主可以将恶意软件插入到着陆页中；点击追踪器可以通过导引到着陆页的最终统一资源定位符 (URL) 的重定向链插入恶意软件；广告主可以将广告或小组件置放在由将恶意软件插入到着陆页上的第三方扩充的页面上；等。

[0007] 一旦知道着陆页具有恶意软件，广告发布者可以排除对着陆页的派发。然而，例如谷歌公司的广告发布者可能访问几亿的广告和与广告相关联的对应的着陆页。如能够被理解的，可能很难深入检查 / 复查每一个着陆页存在恶意软件。

发明内容

[0008] 在此公开的内容是用于检测并处理赞助内容中的恶意软件的装置、方法和系统。在一个实施方式中，识别与赞助内容相关联的着陆页的侵入特征。基于所识别的侵入特征生成着陆页的特征分值，以及如果着陆页的特征分值超过特征阈值，则将着陆页分类为候

选着陆页。可以将候选着陆页提供给恶意软件检测器以确定着陆页中是否存在恶意软件。在一些实施方式中，可以暂停与候选着陆页相关联的赞助者帐户。在一些实施方式中，可以暂停与候选着陆页相关联的广告。

[0009] 在另一个实施方式中，方法包括将与广告相关联的着陆页分割成训练着陆页和测试着陆页。针对训练着陆页的侵入特征迭代地训练分类模型，以及针对测试着陆页的侵入特征迭代地测试分类模型。训练和测试继续直到出现停止事件。响应于停止事件，将从迭代训练和测试取得的特征权重与侵入特征的关联存储在分类模型中。

[0010] 在另一个实施方式中，系统包括评分引擎，其包括存储在计算机可读介质中并且可由处理系统执行的软件指令。在执行时，处理系统识别与赞助内容相关联的着陆页并且识别着陆页的侵入特征。基于所识别的侵入特征生成着陆页的特征分值，以及如果着陆页的特征分值超过特征阈值，则将着陆页分类为候选着陆页。可以将候选着陆页提供给恶意软件检测器以确定着陆页中是否存在恶意软件。在一些实施方式中，可以暂停与候选着陆页相关联的赞助者帐户。在一些实施方式中，可以暂停与候选着陆页相关联的广告。

[0011] 将在附图和下面的描述中阐述本说明中描述的主题的一个或多个实施例的细节。主题的其它特征、方面以及优势从描述、附图以及权利要求将变得显而易见。

附图说明

[0012] 图 1 是示例在线广告系统的框图。

[0013] 图 2 是赞助内容的示例子联合的流程图。

[0014] 图 3 是示例赞助内容处理系统的框图。

[0015] 图 4 是构建分类模型的示例训练过程的框图。

[0016] 图 5 是利用分类模型的另一个示例赞助内容处理系统的框图。

[0017] 图 6 是另一个示例赞助内容处理系统的框图。

[0018] 图 7 是用于为侵入检测识别候选着陆页的示例过程的流程图。

[0019] 图 8 是用于将候选着陆页提交给侵入检测引擎的示例过程的流程图。

[0020] 图 9 是用于处理上诉请求的示例过程的流程图。

[0021] 图 10 是用于处理上诉请求的另一个示例过程的流程图。

[0022] 图 11 是用于生成分类模型的示例过程的流程图。

具体实施方式

[0023] 图 1 是示例在线广告系统 100 的框图。在一些实施方式中，一个或多个广告主 102 可以直接或间接地输入、维护以及追踪广告管理系统 104 中的广告（“ad”）信息。虽然引用的是广告，但是系统 100 可以递送其它形式的内容，包括其它形式的赞助内容。广告可以是以下列形式：诸如横幅广告的图形广告、纯文本广告、图像广告、音频广告、视频广告、组合任何这样的组件的一个或多个的广告等。广告还可以包括嵌入式信息，诸如链接、元信息和 / 或机器可执行的指令。一个或多个发布者 106 可以向系统 104 提交对广告的请求。系统 104 通过将广告发送给发出请求的发布者 106 以在发布者的 web 资产（例如网站或其它网络分布内容）的一个或多个上置放来作出响应。广告可以包括嵌入链接着陆页，例如广告主 102 网站上的页面，当用户点击呈现在发布者网站上的广告时用户被指引到所述着陆

页。

[0024] 诸如用户 108 和广告主 102 的其他实体可以向系统 104 提供使用信息，诸如与广告有关的转换或点进是否已发生。该使用信息可以包括与已派发的广告有关的被测量或观察到的用户行为。系统 104 执行金融交易，诸如基于使用信息向发布者 106 赊账并且向广告主 102 收费。

[0025] 诸如局域网 (LAN)、广域网 (WAN)、因特网或其组合的计算机网络 110 连接广告主 102、系统 104、发布者 106 和用户 108。

[0026] 发布者 106 的一个示例是常规内容服务器，其接收对内容（例如，文章、讨论话题、音乐、视频、图形、搜索结果、网页收录、信息供给等）的请求，并且响应于请求检索所请求的内容。内容服务器可以将对广告的请求提交给系统 104 中的广告服务器。广告请求可以包括所期望的广告的数量。广告请求还可以包括内容请求信息。该信息可以包括内容本身（例如，页面或其它内容文档）、与内容或内容请求相对应的类别（例如，艺术、商业、计算机、艺术电影、艺术音乐等）、内容请求的部分或全部、内容寿命、内容类型（例如，文本、图形、视频、音频、混合媒体等）、地理位置信息等。

[0027] 在一些实施方式中，内容服务器可以将所请求的内容与系统 104 提供的广告中的一个或多个相组合。该组合的内容和广告可被发送给请求内容的用户 108 以在查看器（例如，浏览器或其它内容显示系统）中呈现。内容服务器可以将关于广告的信息传送回广告服务器，所述信息包括描述如何、在何时和 / 或在何处呈现广告（例如，以 HTML 或 JavaScript™) 的信息。

[0028] 另一个示例发布者 106 是搜索服务。搜索服务可以接收对搜索结果的查询。作为响应，搜索服务可以从文档的索引（例如，从网页的索引）检索相关搜索结果。在澳大利亚布里斯班举行的 SeventhInternational World Wide Web Conference（第七届国际万维网会议）上 S.Brin 和 L.Page 发表的论文“The Anatomy of a Large-Scale HypertextualSearch Engine（剖析大规模超文本搜索引擎）”和美国专利 No.6,285,999 中描述了示例性搜索服务。搜索结果可以包括例如网页标题的列表、从那些网页提取的文本的摘录以及指向那些网页的超文本链接，并且可以被分组成预定数量（例如 10) 的搜索结果。

[0029] 搜索服务可以将对广告的请求提交给系统 104。请求可以包括所期望的广告的数量。该数量可以取决于搜索结果、搜索结果占据的屏幕或页面空间量、广告的尺寸和形状等。在一些实施方式中，所期望的广告的数量将是从 1 至 10 或者从 3 至 5。对广告的请求还可以包括（如被输入或被解析的）查询、基于查询的信息（诸如地理位置信息、查询是否来自会员和这样的会员的标识符）和 / 或与搜索结果相关联或者基于搜索结果的信息。这样的信息可以包括例如与搜索结果有关的标识符（例如，文档标识符或“docID”）、与搜索结果有关的分值（例如，信息检索（“IR”) 分值）、从所识别的文档（例如网页）提取的文本的摘录、所识别的文档的全文、所识别的文档的特征向量等。在一些实施方式中，IR 分值可以从例如与查询和文档相对应的特征向量的点积、页面排名分值和 / 或 IR 分值与页面排名分值的组合等来计算。

[0030] 搜索服务可以将搜索结果与系统 104 提供的广告中的一个或多个相组合。然后可以将该组合信息转发给请求内容的用户 108。搜索结果可以被维护为区别于广告，以免用户

在付费广告和推测为中性的搜索结果之间产生混淆。最后，搜索服务可以将与广告有关以及在何时、在何处和 / 或如何呈现广告的信息传送回系统 104。

[0031] 如可从前述所理解的，广告管理系统 104 可以为诸如内容服务器和搜索服务的发布者 106 服务。系统 104 允许对定向到由内容服务器提供的文档的广告进行派发。例如，网络或互联网络可以包括响应于来自具有用于售卖的广告位的搜索服务的请求而派发定向广告的广告服务器。假设互联网络是万维网。搜索服务爬取大量或所有的内容。该内容的一些会包括可用的广告位（也被称为“库存”）。更具体地，一个或多个内容服务器可以包括一个或多个文档。文档可以包括网页、电子邮件、内容、嵌入式信息（例如，嵌入式媒体）、元信息和机器可执行的指令以及可用的广告位。被插入到文档中的广告位中的广告可以在每次派发文档时改变或者替选地可以具有与给定文档的静态关联。

[0032] 在一个实施方式中，广告管理系统 104 可以包括选择来自广告主 102 的广告的拍卖过程。可以允许广告主 102 对广告主乐意为广告的每一次点击而支付的金额，例如在例如用户点击广告时广告主支付的每点击成本金额，进行选择或者出价。每点击成本可以包括最大每点击成本，例如广告主乐意为基于例如查询中的单词或多个单词的关键词的广告的每一次点击支付的最大金额。然而，也可以使用其它出价类型。基于这些出价，广告可以被选择并且排名以供呈现。

[0033] 在一些实施方式中，系统 104 包括广告页恶意软件检测系统，其可以确定赞助内容（例如，广告的着陆页）包含恶意软件的可能性。恶意软件可以包括任何类型的计算机毒害物，诸如不诚实的广告软件、计算机病毒、间谍软件、特洛伊木马、计算机蠕虫或其它这样的恶意的、非选择的和 / 或不需要的软件。具体地，恶意软件可以包括在着陆在诸如广告的着陆页的网页时自动发生的任何可疑的软件安装。在一些实施方式中，广告页恶意软件检测系统可以覆盖用户必须点击页面上的链接（诸如“免费下载”）以安装恶意软件的情况。然而，软件无需进行恶意行为才成为恶意软件；侵入安装的任何软件可以被认为是恶意软件，而不管软件进行的行为是否是恶意的。因此，除特洛伊木马、病毒、蠕虫以及浏览器漏洞外，诸如监视软件、起始页面劫持等的不一定损害计算机系统的其它软件可以被认为是恶意软件。

[0034] 恶意软件检测系统可以例如为恶意软件自动测试着陆页（例如，由嵌入的或与赞助内容相关联的 URL 限定的网页）并且在检测到恶意软件时采取适当的行为。这样的行为可以遵照预定策略，诸如暂停广告主的帐户（例如，关于谷歌 AdSense 或 AdWords 的广告主的帐户）、将与着陆页相关联的广告“标记”为恶意软件相关以及帮助最终用户避免这样的广告在未来的负面影响。恶意软件检测系统可以为广告主提供过程以使其“标记的”广告被重查以及其帐户被解除暂停。此外，如果恶意软件检测系统重查广告主的标记的广告的着陆页并且确定关联的着陆页是洁净的（例如，没有恶意软件的），则可以恢复（或澄清）广告主的帐户。在一些实施方式中，可以完全禁止与着陆页相关联的广告，例如可以排除对广告的派发。

[0035] 在一些实施方式中，恶意软件检测系统可以具有暂停广告群组的灵活性，所述广告群组诸如广告组或广告宣传活动中的所有广告、或者带有共同 URL 的所有广告。例如，恶意软件检测系统可以确定只是广告主的广告的子集包含恶意软件，从而仅暂停这些广告。这样的确定可以基于广告的着陆页共享的共同特征。

[0036] 可能在广告的着陆页或重定向链中遭遇恶意软件，或者恶意软件可以以各种方式引起。具体地，重定向链可以包括一系列 URL，其包括所点击的广告（或目的 URL）、作为点击广告的结果的由脚本等例示的 URL 以及广告的着陆页的最终 URL。在一些情况下，广告主的站点能够被攻击并且恶意软件直接插入到着陆页上。在另一个示例中，恶意广告主可以在其广告着陆页上故意安装或许可恶意软件。在第三个示例中，点击追踪器可以在到达最终 URL 之前的整个重定向链插入恶意软件。在第四个示例中，广告主可以在其可以由插入恶意软件的第三方扩充的着陆页上安装广告和 / 或小组件。在恶意软件的这些和其它示例中，当用户点击这样的广告时，用户的计算机能够被侵入软件的安装危害。

[0037] 例如，图 2 是赞助内容的示例子联合的流程图。考虑在受欢迎的网站 204 上的广告 202。例如，讨论中的网站或网页可以包括来自有信誉的广告公司 206a 的横幅广告。用户可以点击例如在用户的家用计算机、PDA 等上执行的 web 浏览器中的广告 202。广告 202 的目的 URL 可以指向或启动引用在第一地理区域 208 的广告公司 206a 的单行脚本（例如，第一 JavaScript）。第一 JavaScript 进而可以生成第二 JavaScript 来重定向至广告公司 206b（例如，另一个广告公司）的 URL。第二 JavaScript 进而可以生成指向可以例如为其广告使用地理定向的另一个广告公司 206c 的更多 JavaScript。地理定向的广告可以导致指向在另一个地理区域 210 的广告公司 206d 的包含 iFrame 的单行 HTML。

[0038] 当试图检索该 iFrame 时，可以将浏览器诸如经由位置头部朝向漏洞服务器的 IP 地址重定向。例如，IP 地址可以是以例如 xx. xx. xx. xx/< 漏洞服务器 >/ 的格式，诸如漏洞服务器 212 的 IP 地址。所提供的 IP 地址可以包括可以使漏洞服务器 212 能够针对用户的浏览器尝试多个漏洞的加密的 JavaScript。结果，可以在用户的计算机上安装数个恶意软件二进制文件。在这种情况下遭遇和 / 或安装的恶意软件对于最初的广告公司 206a 而言可能是不知道的。然而，从目的 URL（例如，广告公司 206a）到与赞助内容相关联的着陆页（例如，在漏洞服务器 212 上）的每一次重定向可以引起另一方控制原始网页上的内容。以这种方式，在此由数个 URL 重定向表征的赞助内容的子联合可以导引用户与恶意软件的非期望遭遇。

[0039] 检测恶意软件可以包括使用商业可获得恶意软件检测软件或其它这样的病毒扫描软件或系统。还可以通过监视系统行为，诸如监视在访问 URL 后对注册表和系统文件的使用，来检测恶意软件。例如，侵入检测引擎可以监视虚拟机上的浏览器的行为以确定是否存在恶意软件。

[0040] 图 3 是示例赞助内容处理系统 300 的框图。与赞助内容有关的数据可以被存储在广告数据库 302 中。例如，广告数据库 302 可以包含数种形式的赞助内容，诸如图形广告、横幅广告、纯文本广告、图像广告、音频广告、视频广告、组合任何这样的组件的一个或多个的广告等。广告可以例如按关键词、广告宣传活动、URL 或其它内容来组织。

[0041] 系统 300 包括可以被用来检测与广告相关联的着陆页或广告自身中的恶意软件的恶意软件评估器 304。例如，恶意软件评估器 304 初始可以为广告的着陆页包括恶意软件的可能性而评估广告的着陆页，并且如果认为着陆页可能包括恶意软件，则恶意软件评估器 304 可以将广告提交给更全面的评估过程。这样的两步评估过程可以导致通过仅对被认为最可能包括恶意软件的候选广告使用更全面的恶意软件评估过程而获得的效率。

[0042] 恶意软件评估器 304 执行的初始评估可以识别广告的着陆页或重定向链中的 URL

的侵入特征。评估可以审查广告的 iFrame 特征、URL 特征、脚本特征等并且将这样的特征与已知与包括恶意软件的着陆页相关联的特征的储存库进行比较。作为对广告的着陆页特征的初始评估的结果，恶意软件评估器 304 可以生成指示广告的着陆页包括恶意软件的可能性的特征分值。例如，更高的分值可以表示广告的着陆页更可能包括恶意软件。可以将具有超过特征阈值的特征分值的任何广告的着陆页分类为用于更全面的恶意软件评估过程的候选。以这种方式，对特征的识别可以促进减少试探，允许系统将着陆页的数量显著减少至随后可以由更全面的恶意软件评估过程评估的候选着陆页的更小集合。

[0043] 在一些实施方式中，恶意软件评估器 304 可以使用实现更全面的恶意软件评估过程的侵入检测引擎 305。例如，恶意软件评估器 304 可以向侵入检测引擎 305 提供网页（例如，广告的着陆页）并且接收关于该网页的侵入分值。在其它实施方式中，恶意软件评估器 304 可以包括侵入检测引擎 305。

[0044] 更全面的过程可以由恶意软件评估器 304 在恶意软件评估器将候选着陆页提交给侵入检测引擎 305 时启动。侵入检测引擎 305 可以包括例如虚拟机，经由该虚拟机，系统 300 可以在浏览器中加载广告、（例如，经由一个或多个 URL 重定向）导航到广告的着陆页以及执行诸如商业可获得计算机恶意软件和病毒检测系统的一个或多个恶意软件检测系统。在该过程期间，虚拟机还可以例如监视对系统文件的使用和对未授权的过程的创建。侵入检测引擎 305 可以生成侵入分值并且将侵入分值提供给恶意软件评估器 304。侵入分值可以指示在广告的着陆页中的恶意软件的层级。如果侵入分值显著高，诸如高于预定义的侵入阈值，则系统 300 可以（例如，在广告数据库 302 中）将广告标记为可能在其着陆页中包含恶意软件。

[0045] 可以排除在广告数据库 302 中标记的广告以不派发给用户，或者可以以某种方式注释广告以指示其的广告的着陆页包括恶意软件的可能性。在一些实施方式中，注释可以包括对广告为恶意软件相关的可能性中的每一个评级的侵入分值。作为确定广告主的赞助内容的任何部分（例如，单个广告着陆页）包括恶意软件的结果，系统 300 可以标记广告主的广告中的一些或全部。系统 300 还可以暂停广告主的帐户，诸如以阻止广告主提交新的广告。诸如在清楚广告为恶意软件相关时，例如相对高的侵入分值，系统 300 可以自动执行一些行为。其它行为可以基于诸如在查阅恶意软件评估的结果后的用户决定。

[0046] 帐户管理器 306 可以从恶意软件评估器 304 接收恶意软件评估的结果。评估可以包括例如赞助者的帐户信息、目的和着陆页以及重定向链中的任何页面的 URL。评估还可以包括识别恶意软件评估器 304 将广告识别为恶意软件相关的原因的信息。帐户管理器 306 的用户可以能够便于基于评估来人工处置广告和 / 或帐户。例如，如果广告主的广告着陆页中的一个或多个被发现包括恶意软件，则用户可以能够暂停关于该广告主的帐户。在另一个示例中，用户可以决定标记广告主的广告宣传活动中的一个或多个广告。

[0047] 在系统 300 内可以存在允许广告主为标记的广告启动上诉过程的客户服务代表 (CSR) 前端 308。例如，客户（例如，广告主）可以具有与恶意软件评估器 304 确定包括恶意软件的赞助内容相对应的一个或多个着陆页。例如，在对这样的站点清除恶意软件后，广告主可以启动对广告的上诉。这样的上诉可以是例如以在 CSR 前端 308 和恶意软件评估器 304 和 / 或帐户管理器 306 之间的通信的方式。通信可以包括例如广告主的名称和待由恶意软件评估器 304 重新评估的着陆页的 URL。如果广告主的对标记的广告的上诉是成功的，

则系统 300 可以对广告解除标记。在一些实施方式中，作为成功上诉的结果，系统 300 还可以恢复广告主的帐户。在一些实施方式中，在广告主上诉广告时，系统 300 可以检查关于该广告主的所有广告并且如果该广告主的广告中的所有广告是洁净的，则才恢复广告主的帐户（以及对广告解除标记）。

[0048] 在一些实施方式中，系统 300 可以包括分层级的暂停帐户模型。例如，基于着陆页中存在恶意软件的可能性，着陆页可以被归类入恶意软件侵入的各种类别或层级中。这样的类别可以包括例如“OK”（例如，确定可能是无恶意软件的）、“可疑的”（例如，可能包含恶意软件）或“确认的”（例如，非常可能或肯定包含恶意软件）。可疑类别可以诸如以基于侵入分值的评级来进一步归类。

[0049] 在一些实施方式中，恶意软件检测分值可以关于帐户累积，以及帐户自身可以被“分层级”入风险类别，风险类别的每一个被不同地处理，在自动查阅、人工查阅以及自动暂停内变化。例如，系统 300 可以在一个或多个广告在“确认的”恶意软件类别中时自动暂停帐户、或者可以在 5％或更多的广告是“可疑的”时暂停帐户等。

[0050] 例如，在一个实施方式中，恶意软件评估器 308 可以识别与具有超过特征阈值的特征分值的赞助者帐户相关联的着陆页。可以对这些着陆页的特征分值进行累积以获取帐户分值，以及可以基于帐户分值将风险类别分配给赞助者帐户。可以基于风险类别来选择用于赞助者帐户的数个帐户补救过程中的一个，例如自动查阅、人工查阅、自动暂停、只是候选着陆页的部分暂停等。

[0051] 对潜在恶意软件的检测可以连续地、周期性地或不定期地发生。例如，广告数据库 302 可以由恶意软件评估器 304 连续地检查。在另一个示例中，广告数据库 302 可以由恶意软件评估器周期性地检查，例如每月或每周检查。在又一个示例中，添加到广告数据库 302 的每一个广告可以在广告被添加到广告数据库 302 时被检查。还可以使用其它检测调度。

[0052] 图 4 是构建分类模型 402 的示例训练过程 400 的框图。分类模型 402 可以被用于评估与广告相关联的着陆页中的特征，诸如可以指示存在恶意软件的可能性的特征，例如小 iFrame、混淆脚本等。在一些实施方式中，可以在训练过程期间为特征分配权重。这样的基于特征的评估可以被用来减少待使用更鲁棒的评估过程来评估的 URL 的数量，所述过程诸如由侵入检测引擎 305 实现的过程。

[0053] 训练过程 400 可以被用来使用“训练”着陆页内容的侵入特征来迭代训练分类模型 402。同时，过程 400 可以使用“测试”着陆页内容的侵入特征来迭代测试分类模型 402。迭代过程 400 可以继续直到出现测试停止事件，诸如在特征权重和侵入特征之间的关联是稳定的确定。这样的确定可以例如通过实现基于线性回归的模型来进行。

[0054] 在用于产生分类模型 402 的训练过程 400 的示例一般流程中，处理可以以对广告 302 的使用开始。可以从着陆页和 URL 404 识别用于训练过程 400 的信息。过程 400 可以将着陆页和 URL 404 进一步分割成“训练”着陆页和“测试”着陆页。例如，较大数量的着陆页（例如 10,000）可以被用作训练示例来训练分类模型 402，而较小数量（例如 1,000）的着陆页可以被用来测试分类模型 402。

[0055] 特征提取引擎 406 可以从着陆页和 URL 404 提取特征。特征可以例如指示与广告相关联的着陆页包括恶意软件的可能性。例如，一个或多个恶意软件相关的（或侵入）特征可以与可以指示将其它 HTML 文档（例如，恶意软件相关的）嵌入到主文档内的企图

的小 iFrame 相对应。侵入特征的另一个示例是不良或可疑的 URL，诸如与恶意软件感染的域的已知列表上的 URL 相匹配的 URL。侵入特征的第三个示例是可疑的脚本语言。例如，JavaScript 或其它脚本语言可以具有已知在派发恶意软件时使用的某些函数调用或语言元素。可以存在数种其它类型的侵入特征，诸如存在多个帧、脚本或 iFrame 在异常位置（例如，在 HTML 的末尾之后）出现，或训练过程 400 随着时间的推移确定的任何其它特征为可能的恶意软件侵入的标志。

[0056] 在一些实施方式中，特征提取引擎 406 可以包括被加权的特征的列表。例如，关于已知恶意软件站点的 URL 的特定侵入特征可以接收比较少可能与恶意软件相关联的侵入特征更高的权重。由于分类模型 402 被用来根据着陆页包括恶意软件的可能性将着陆页分类，所以可以随着时间的推移调整特征的权重。

[0057] 权重可以是累积的，使得着陆页包括恶意软件的总体可能性可以通过将与检测到的特征相对应的权重进行相加或组合来确定。在一些实施方式中，对于可以在着陆页中检测到的对应的特征的每一次出现，可以将特征的权重包括在总和中。在其它实施例中，可以将特征的权重添加到总分值一次，而不考虑特征在广告中出现的次数。还可以使用基于特征权重的其它评估。

[0058] 虽然许多特征可以具有对应的正权重，但是其它特征可以具有负权重。例如，特征 A（例如，对应于可能的恶意软件相关的函数调用）可以具有权重 2.5。同时，特征 X 的存在可以部分否认特征 A 是恶意的可能性，促使系统 400 将负权重分配给特征 X。

[0059] 在训练过程 400 的训练阶段中可以使用控制评估 408。控制评估 408 可以包括广告着陆页的人类评估。例如，对用于特定广告的着陆页的人类查阅可以包括对广告的特征的查验。查阅还可以提供着陆页的包括恶意软件的可能性的总体评级，诸如极度恶意软件感染、部分恶意软件感染等。

[0060] 由控制评估 408 生成的信息可以在将特征权重分配给由特征提取引擎 406 提取的特征的训练阶段期间引用。例如，机器学习引擎 410 可以例如通过查验在其它 URL（例如，来自“测试”着陆页的 URL）中的类似特征，将特征权重分配给特征以测试控制评估 408 的结果。具体地，机器学习引擎 410 可以使用来自测试着陆页的特征来迭代精化特征权重与侵入特征的关联。

[0061] 这样的精化可以例如由基于线性回归的模型来实现。例如，机器学习引擎 410 可以使用以着陆页和 URL 404 分割的训练和测试着陆页。机器学习引擎 410 可以例如基于训练和测试着陆页来调整特征权重以为测试着陆页生成特征分值。如果特征分值产生接近控制评估结果的恶意软件检测结果，则分类模型可以被认为是训练的。相反，如果特征分值产生与控制评估结果极大不同的恶意软件检测结果，则机器学习引擎 410 可以重新调整特征权重。例如，经过数次迭代，机器学习引擎 410 可以确定特征 X 被加权得太高，因此可以减少与特征 X 相关联的特征权重。

[0062] 针对训练和测试着陆页的侵入特征对分类模型 402 的迭代训练和测试可以继续直到出现测试停止事件，例如将测试结果收敛到控制评估 408，或直到达到迭代界限。在停止事件后，特征权重与侵入特征的关联可以在分类模型 402 中持久留存。

[0063] 还可以使用训练分类模型 402 的其它过程。

[0064] 图 5 是利用分类模型 402 的另一个示例赞助内容处理系统 500 的框图。系统 500

包括使用分类模型402对来自广告数据库504的广告进行评分的评分引擎502。例如，使用存储在分类模型402中的特征权重，评分引擎502可以对所处理的来自广告数据库504的广告的着陆页的特征进行评分。评分高于预定阈值的任何广告可以被识别为候选URL 506。

[0065] 候选URL 506可以包括恶意软件评估器508全面查验可能所需的与广告相关联的信息。例如，候选URL 506可以包括广告的URL和提供赞助内容的广告主的帐户信息。广告的URL（或用于广告的一些其它标识符）可以例如被用来识别恶意软件评估器506可能所需的关于广告数据库504中的广告的额外信息。广告的URL还可以由恶意软件评估器506用来模拟在用户的浏览器中对广告的选择。例如，恶意软件评估器506可以将着陆页提供给侵入检测引擎305，其可以将URL加载入包括病毒检测软件并且监视对系统文件的使用以及未授权的过程的创建的虚拟机。

[0066] 在一些实施方式中，在恶意软件评估器508（例如，基于从侵入检测引擎305接收的高侵入分值）确定候选URL感染了恶意软件时，可以对其它相关的候选URL 506分配类似的分值。例如，可以清楚的是，具有相同域名的候选URL 506也同样有可能被感染。这样的确定可以部分基于地理因素，例如如果域来自统计上已知具有较高的受感染域率的俄罗斯、中国或任何其它国家。

[0067] 图6是另一个示例赞助内容处理系统600的框图。系统600包括可以检测恶意软件与广告数据库604中的广告相关联的可能性的广告恶意软件检测系统602。广告恶意软件检测系统602还可以促进上诉过程，通过该上诉过程广告主可以请求对已被标记为与恶意软件相关联的广告的重新评估。在一些实施方式中，广告恶意软件检测系统602可以包括软件指令，其连续执行来例如使用来自广告604的信息以持续识别广告的着陆页中的恶意软件。例如，识别过程可以涉及监视系统行为，诸如在用户访问URL后监视对注册表和系统文件的使用。在另一个示例中，识别过程可以涉及对每一个广告主的着陆页URL的定期查验、或可以涉及被认为可能包含恶意软件的一个或相应着陆页URL。这样的过程可以监视可以指示存在恶意软件的可能性的特定广告着陆页特征。在其它实施方式中，广告恶意软件检测系统602的一个或多个组件可以在强力(brute force)过程中用来爬取广告数据库604以为与恶意软件的可能的关联查验着陆页URL。对广告具有关联的恶意软件的确定可以基于单个广告、广告群组、关键词、一个或多个相关URL、广告主的帐户内的广告群组或以上的一些组合。

[0068] 在一个实施方式中，可以将来自广告数据库604的信息提供给广告组标准特征数据库606和URL特征数据库608。例如，数据库606和608中的信息可以包括来自广告的相关信息，诸如URL、来自广告的关键词、关联的广告主的名称、广告主的帐户信息等。对该信息的提供可以例如消除对存储图像、视频、音频或其它这样的广告相关信息的需要。使广告信息本地于广告组标准特征数据库606和URL特征数据库608还可以提供组织和/或索引数据的优势以在广告恶意软件检测系统602内更有效使用。存储在数据库606和608中的这样的信息可以在不必爬取广告的着陆页的情况下足以确定与特定广告的基于恶意软件特征的关联。在另一个实施方式中，系统600可以爬取广告的着陆页并且使用从着陆页可获取的信息，而不是使用数据库606和608（或除使用数据库606和608之外）。

[0069] 广告组标准特征数据库606可以包含广告主的一个或多个广告组的信息、与广告相关联的关键词、产品归类信息、广告主的帐户信息以及由广告恶意软件检测系统602使

用的其它广告相关信息。URL 特征数据库 608 可以包含每一个单个广告的 URL（例如，着陆页 URL）、广告主的名称以及可以允许广告组标准特征数据库 606 中的关联数据被访问的任何其它信息或索引。

[0070] 广告恶意软件检测系统 602 包括取样器 610，其可以在识别可能包含恶意软件的广告时被用作粗过滤器。具体地，取样器 610 可以识别为其推荐恶意软件检测的广告。识别过程可以使用存储在广告组标准特征数据库 606 和 URL 特征数据库 608 中的广告相关信息。例如，取样器 610 可以为在广告组标准特征数据库 606 中识别的预定广告内容特征的集合中的任何内容特征搜索广告。

[0071] 取样器 610 可以使用关于图 4 描述的分类模型 402。例如，取样器 610 可以将所处理的来自广告数据库 606 和 608 的广告的特征与在分类模型 402 中表示的加权特征进行比较。基于广告的着陆页中的一个或多个特征的累积或组合特征权重，取样器 610 可以确定广告的着陆页超过特征阈值。如此，广告的 URL 可以被认为是用于更全面的恶意软件检测的候选 URL。

[0072] 在一些实施方式中，URL 特征数据库 608 可以包括取样器 610 中的混淆检测器可以使用来为混淆脚本审查 HTML 页面的混淆信息，所述混淆脚本诸如以 JavaScript、VBScript 等编写的脚本。这样的脚本通常可以包含字符的显然是混乱数据的集合，其在用户点击广告时将自身重写为另一个 URL 字符串，然后再次重写为又一个字符串等直到漏洞代码被编写或下载到计算机设备上。可以循着重定向链发生的该层级或重写能够使识别恶意的 HTML 代码变得困难。

[0073] 在一些实施方式中，URL 特征数据库 608 可以包括地理位置信息。这样的信息可以被用来例如将用于广告的 URL 地理地归类。通常可能从某些国家提供恶意软件，因此分析嵌入式链接的位置信息可以帮助识别潜在的恶意软件站点。例如，具有指向在地理遥远的位置的以恶意软件的高发生率而闻名的站点的 iFrame 的 US-.com 域可以提供潜在恶意软件的强信号。

[0074] 在取样器 610 识别了被怀疑包含恶意软件的候选广告时，取样器 610 可以将候选 URL 和帐户信息发送给恶意软件中心 612。恶意软件中心 612 可以被用作中央接口，其用于接收待被更全面地检查恶意软件的广告，以及如将在下面描述的，用于接收对被标记为包含恶意软件的广告的上诉。对于取样器 610 请求恶意软件中心 612 查阅的任何广告，恶意软件中心 612 可以以广告的 URL 和诸如与广告相关联的广告主的帐户信息的对应的追踪信息更新状态数据库 614。在一些实施方式中，存储在状态数据库 614 中的信息可以包括取样器 614 认为用于更高级的恶意软件检测的原因的信息。在一些实施方式中，原因可以被用来对在状态数据库 614 中的广告状态分组，以便将广告状态分组用于更有效的处理。

[0075] 在一些实施方式中，取样器 610 还可以评估域以及用于那些域的 URL（或指向那些域的链接）的相对寿命。由于恶意软件通常分发自新的站点，所以域的寿命可以被用来识别可疑的恶意软件站点。例如，新的分发站点被不断创建并且可以在站点被撤下前仅存在数周。为了确定域的寿命，取样器 610 可以使用可能可用的最近激活的域名的公共或私人列表，例如来自域登记清算机构。

[0076] 在一些实施方式中，恶意软件中心 612 可以被用作用于从其它广告管理系统 104 接收广告恶意软件检测请求的中央接口。例如，虽然广告恶意软件检测系统 602 可以是谷

歌的 AdSense 系统的组件，但是竞争广告管理系统 104 可以支付费用以使广告在其控制下进行恶意软件审查。如此，广告恶意软件检测系统 602 可以被用作用于数个广告管理系统 104 的恶意软件检测的清算机构。

[0077] 恶意软件检测器 616 可以处理状态数据库 614 中的条目所表示的广告。例如，恶意软件检测器 616 可以使用关于每一个广告的 URL 和帐户 ID 来处理一个或多个广告。如果需要关于广告的额外信息（例如，未被存储在状态数据库 614 中），则恶意软件检测器 616 可以从广告数据库 604 拉取关于广告的额外信息。这样的信息可以包括例如帐户信息或对于初始粗过滤器审查可能没有提供给取样器 610 的广告自身的部分。

[0078] 恶意软件检测器 616 然后可以促使更全面的审查被执行。另外，恶意软件检测器 616 可以将 URL 提交给例如侵入检测引擎 305 的执行更详细的恶意软件评估的侵入检测引擎，所述更详细的恶意软件评估诸如仔细查验“目的”URL、“最终”URL、重定向链中的 URL 以及（例如，最终 URL 所识别的）广告的着陆页。

[0079] 恶意软件检测器 616 可以从侵入检测引擎 305 接收关于目的 URL 的侵入分值。对于具有高于预定阈值的侵入分值的每一个着陆页，广告恶意软件检测系统 602 可以采取一个或多个预定行为，诸如自动将广告标记为恶意软件相关并且暂停广告主的帐户，或将这样的信息提供给可以人工暂停恶意广告主的帐户和 / 或排除其广告的用户。可以将恶意软件检测器 616 可以应用的侵入分值阈值保守设置得很高，以免产生重大假阳性。

[0080] 在一些实施方式中，可以利用恶意软件检测器 305 实现侵入检测引擎或将其与恶意软件检测器 305 集成在一起。

[0081] 在检测到广告恶意软件时出现的行为可以遵照预定策略。例如可以人工暂停广告主的帐户，并且可以通知广告主。可以对与恶意软件相关联的广告进行标记以避免向用户派发该广告。恶意软件检测器 616 可以将关于标记的广告、暂停的帐户等的信息提供给状态数据库 614。在一些实施方式中，过程可以定期运行以使用状态数据库 614 中的这样的信息来更新广告数据库 604。

[0082] 客户前端 618 可以被用作例如客户服务代表的用于用户的图形用户界面 (GUI)，以查阅恶意软件检测器 616 所执行的广告恶意软件检测的任何结果。例如，结果可以列出具体着陆页的实例以及确定其包含恶意软件的原因。可以将实例以各种方式分组或分类，诸如按广告主帐户、URL 等。

[0083] 上诉过程可以允许具有被标记的广告的广告主使广告被广告恶意软件检测系统 602 重查。例如，广告主可以在被通知广告的着陆页包含恶意软件后去除恶意软件的广告的最终 URL、或重定向链中的所有 URL，并且然后联系客户服务代表作为上诉过程的一部分。客户服务代表可以利用客户前端 618 来将上诉请求发送给恶意软件中心 612。每一个上诉请求可以表示广告主请求广告恶意软件检测系统 602 重新评估恶意软件内容的一个或多个广告。例如，如果广告恶意软件检测系统 602 先前将广告主的广告标记为恶意软件相关，并且广告主已清除与该广告相关联的着陆页 URL，则请求可以是重新评估该特定的广告。

[0084] 恶意软件中心 612 可以接收上诉请求并且更新上诉数据库 620。具体地，可以将未决的和完成的上诉请求存储在上诉数据库 620 中。存储在上诉数据库 620 中的关于每一个广告的信息可以包括例如广告主名称、广告主的帐户信息、与广告的着陆页相关联的 URL 和重定向链中的 URL、以及可以被用来处理上诉的任何其它信息。

[0085] 为了作为上诉处理，恶意软件检测器 616 可以使用类似于为恶意软件初始评估广告的着陆页的上述过程的过程。在一些实施方式中，上诉过程还可以自动包括对广告主的所有广告、广告组或可以被用来搜索广告主可能具有的其它恶意软件相关的广告的任何其它这样的组中的所有广告的着陆页重新评估。

[0086] 在处理上诉时，恶意软件检测器 616 可以使用存储在上诉数据库 620 中的关于每一个广告的信息。恶意软件检测器 616 可以使用与上述过程类似的过程来评估广告的着陆页、生成侵入分值并且应用阈值来确定广告的着陆页是否可能已清除了恶意软件。可以将广告着陆页重新评估的结果存储在上诉数据库 620 中。在一些实施方式中，过程可以定期运行以使用上诉数据库 620 中的这样的信息来更新广告数据库 604。

[0087] 在恶意软件上诉的一个示例情况中，客户可以接收陈述客户的帐户因为恶意软件而被暂停的通知，诸如电子邮件。通知可以包括找到恶意软件的位置的细节（例如，目的 URL、帐户信息等）。通知还可以提供对如何移除恶意软件的建议，并且可以指导例如与恶意软件客户支持代表的后续操作。客户然后可以清除其与恶意软件相关联的着陆页和 / 或其它 URL，并且使用客户前端 618 来启动上诉过程。如果恶意软件检测器 616 确定广告的着陆页现在没有恶意软件，则客户可以接收上诉成功并且现恢复帐户的通知。然而，如果恶意软件检测器 616 确定广告的着陆页仍然包括恶意软件，则客户可以接收上诉被拒绝的通知，其包括关于检测到的恶意软件的详细信息。在一些实施方式中，可以按组完成关于恶意软件检测和上诉结果的通知过程，例如不会使客户对大量的电子邮件通知不知所措。

[0088] 在一些实施方式中，以每个广告为基础来排除与赞助者帐户相关联的广告，例如仅排除具有超过侵入阈值的侵入分值的广告以不被派发。在上诉时，将候选着陆页重新提交给侵入检测引擎，并且从侵入检测引擎接收候选着陆页的另一个侵入分值。广告依赖在上诉期间接收的侵入分值来保持暂停或被恢复。

[0089] 在一些实施方式中，如果确定帐户中的任何一个广告具有超过侵入阈值的侵入分值，则以每个帐户为基础来排除与赞助者帐户相关联的广告。在上诉时，为恶意软件识别并检查赞助者帐户中的所有广告。如果确定与赞助者帐户相关联的着陆页中的任何一个具有超过侵入阈值的侵入分值，则帐户保持暂停。

[0090] 图 7 是用于识别用于侵入检测的候选着陆页的示例过程 700 的流程图。例如，过程 700 可以使用存储在计算机可读介质中并且由处理系统执行的软件指令来实现。由过程 700 识别的候选着陆页可以由评分引擎 502（参见图 5）识别以及由恶意软件检测器 508 使用。这样的候选页可以是比过程 700 用来识别候选着陆页的着陆页的总集合显著小的页面数量。

[0091] 阶段 702 识别与赞助内容相关联的着陆页。例如，着陆页可以是用户在点击广告后可以在 web 浏览器中看到的关于广告的着陆页。总的来说，“着陆页”的语境可以包括任何内容或头部，包括 web 浏览器的用户在广告点击后可以遭遇或看到的重定向。

[0092] 阶段 704 识别着陆页的侵入特征。例如，过程 700 可以使用图 5 中的评分引擎 502 来识别着陆页特征，诸如一个或多个 iFrame 特征、一个或多个 URL 特征和 / 或一个或多个脚本特征。在另一个示例中，参考图 6 描述的取样器 610 可以从广告组标准特征数据库 606 和 URL 特征数据库 608 识别特征。

[0093] 阶段 706 基于所识别的侵入特征生成关于着陆页的特征分值。例如，评分引擎

502(参见图 5)可以使用来自分类模型 402 的加权分值来为来自广告数据库 504 的广告的着陆页生成特征分值。在另一个示例中,取样器 610 可以基于使用自特征数据库 606 和 URL 特征数据库 608 的来自广告的着陆页的特征来生成特征分值。

[0094] 阶段 708 确定关于着陆页的特征分值是否超过特征阈值。例如,评分引擎 502 可以确定为广告的着陆页生成的特征分值是否超过预定的特征阈值。在另一个示例中,取样器 610 可以确定为广告生成的特征分值是否超过预定的特征阈值。例如,特征阈值可以是数字。在一些实施方式中,对于不同层级的广告主可以存在不同的特征阈值,所述层级诸如基于恶意软件风险的层级。例如,已知具有很少或没有恶意软件相关的广告的广告主可以具有较高的阈值;或广告主可以请求建立更低的阈值,以便更容易地识别潜在的受感染的广告以防止差的客户体验;等。

[0095] 如果关于着陆页的特征分值超过特征阈值,则阶段 710 将着陆页分类为候选着陆页。例如,如果评分引擎 502 确定为广告的着陆页生成的特征分值超过预定的特征阈值,则评分引擎 502 可以输出对应的候选 URL 506。在另一个示例中,如果取样器 610 确定为广告的着陆页生成的特征分值超过预定的特征阈值,则取样器 610 可以将候选 URL 提供给恶意软件中心 612。

[0096] 图 8 是用于将候选着陆页提交给侵入检测引擎的示例过程 800 的流程图。例如,由过程 800 提交的候选着陆页可以由过程 700 识别。过程 800 可以使用存储在计算机可读介质中并且由处理系统执行的软件指令来实现。

[0097] 阶段 802 将候选着陆页提交给侵入检测引擎。例如,参考图 5,系统 500 可以将候选 URL 506 提供给恶意软件评估器 508,其可以将候选 URL 506 提供给侵入检测引擎 305。在另一个示例中,可以将状态数据库 614 中表示的候选 URL 提供给恶意软件检测器 616(参见图 6)。

[0098] 阶段 804 从侵入检测引擎接收关于候选着陆页的侵入分值。例如,参考图 6,状态数据库 614 可以从恶意软件检测器 616 接收侵入分值。侵入分值可以与恶意软件检测器 616 处理的来自状态数据库 614 的广告的着陆页相对应。

[0099] 如果侵入分值超过侵入阈值,则阶段 806 排除对与候选着陆页相关联的广告的派发。例如,如果由恶意软件检测器 616 处理的广告的着陆页的侵入分值超过侵入阈值,则恶意软件检测器 616 可以以对应的广告待被标记的信息来更新状态数据库 614。可以将状态数据库 614 中的这样的信息稍后用来更新广告数据库 604。可以以各种方式排除在广告数据库 604 中被标记的广告,诸如通过将已派发的广告(例如,在用户的浏览器中)标记为包含潜在的恶意软件或通过排除广告被派发。阶段 806 中的排除还可以包括暂停广告主的帐户,或在层级帐户系统中,提高关于广告主的恶意软件风险评级。

[0100] 图 9 是用于处理上诉请求的示例过程 900 的流程图。例如,上诉请求可以由广告主在诸如通过过程 800 排除该广告主的广告中的一个或多个后提出。过程 900 可以使用存储在计算机可读介质中并且由处理系统执行的软件指令来实现。

[0101] 阶段 902 接收对赞助者帐户的上诉请求。例如,上诉可以源自图 6 的客户前端 618。上诉请求可以例如由可以将关于上诉请求的信息存储在上诉数据库 620 中的恶意软件中心 612 接收。

[0102] 阶段 904 将候选着陆页重新提交给侵入检测引擎。例如,系统 600 可以使用存储在

上诉数据库 620 中的对应于上诉的信息来将候选着陆页重新提交给恶意软件检测器 616，其可以包括侵入检测引擎或与侵入检测引擎通信。

[0103] 阶段 906 从侵入检测引擎接收候选着陆页的另一个侵入分值。例如，作为阶段 904 的重新提交的结果，可以生成并接收关于广告的新的侵入分值。总的来说，例如，如果对广告上诉的广告主自此已去除了恶意软件的广告的着陆页或例如通过雇用新的发布者为广告提供了新的着陆页，则该侵入分值对于广告的着陆页可以是较低的。

[0104] 阶段 908 确定侵入分值是否超过侵入阈值。如果侵入分值超过侵入阈值，如果另一个侵入分值超过侵入阈值，则阶段 910 排除对与候选着陆页相关联的广告的派发。例如，如果由恶意软件检测器 616 处理的广告的着陆页的新的侵入分值超过侵入阈值，则恶意软件检测器 616 可以以对应的广告仍然与恶意软件相关联的信息更新上诉数据库 620。

[0105] 如果侵入分值没有超过侵入阈值，如果另一个侵入分值没有超过侵入阈值，则阶段 912 允许对与候选着陆页相关联的广告的派发。例如，如果由恶意软件检测器 616 处理的广告的着陆页的新的侵入分值没有超过侵入阈值，则恶意软件检测器 616 可以以对应的广告现为洁净的并且可以在没有任何限制的情况下被派发的信息更新上诉数据库 620。

[0106] 图 10 是用于处理上诉请求的另一个示例过程 1000 的流程图。例如，上诉请求可以由广告主在诸如通过过程 800 排除该广告主的广告中的一个或多个后提出。过程 1000 可以使用存储在计算机可读介质中并且由处理系统执行的软件指令来实现。

[0107] 阶段 1002 识别与广告相关联的赞助者帐户，赞助者帐户包括额外的广告。例如，参考图 6，赞助者帐户可以与恶意软件检测器 616 确定广告的着陆页感染了恶意软件的广告相关联。例如，可以在广告组标准特征数据库 606 中为 URL 特征数据库 608 识别帐户。

[0108] 如果候选着陆页的侵入分值超过侵入阈值，则阶段 1004 排除对与赞助者帐户相关联的额外广告的派发。例如，使用在阶段 1002 中识别的赞助者的帐户信息，恶意软件检测器 616 可以排除对广告主的额外广告的派发。特别地，根据过程 1000 所表示的业务策略，一旦确定广告主的一个广告与恶意软件相关联，则可以标记（并且排除）该广告主的该广告和所有其它广告。

[0109] 阶段 1006 接收对赞助者帐户的上诉请求。例如，上诉可以源自执行客户前端 618（参见图 6）的用户。上诉请求可以例如由可以将关于上诉请求的信息存储在上诉数据库 620 中的恶意软件中心 612 接收。

[0110] 阶段 1008 将候选着陆页和与额外广告相关联的额外着陆页提交给侵入检测引擎。例如，系统 600 可以使用存储在上诉数据库 620 中的对应于上诉的信息来将广告主的候选着陆页中的全部候选着陆页提交给恶意软件检测器 616，其可以包括侵入检测引擎或将着陆页信息提供给侵入检测引擎。作为过程的一部分，与候选着陆页相对应的帐户信息可以被用来识别广告数据库 604 中的与广告主的帐户相对应的其它广告。具体地，候选着陆页可以包括原始候选着陆页和与广告主的额外广告相关联的额外着陆页。

[0111] 阶段 1010 从侵入检测引擎接收候选着陆页的另一个侵入分值和关于额外着陆页的额外侵入分值。例如，作为恶意软件检测器 616 评估广告主的全部候选着陆页的结果，可以生成与着陆页相对应的侵入分值。特别地，侵入分值可以被存储在上诉数据库 620 中（或由上诉数据库 620 接收）。在一些实施方式中，可以将额外着陆页的侵入分值存储在状态数据库 614 中。

[0112] 阶段 1012 确定关于着陆页的侵入分值是否超过侵入阈值。例如，如果存在的话，则恶意软件检测器 616 可以确定在阶段 1010 所接收的着陆页的侵入分值中的哪些超过侵入阈值。

[0113] 如果着陆页中的任何着陆页的侵入分值超过侵入阈值，则阶段 1014 排除对与赞助者帐户相关联的广告的派发。例如，如果恶意软件检测器 616 确定侵入分值中的任何侵入分值超过侵入阈值，则恶意软件检测器 616 可以以赞助者的广告（作为整体）仍然包括恶意软件并且可以排除被派发的信息更新上诉数据库 620。

[0114] 图 11 是用于生成分类模型的示例过程 1100 的流程图。例如，过程 1100 可以被用来生成分类模型 402。过程 1100 可以使用存储在计算机可读介质中并且由处理系统执行的软件指令来实现。

[0115] 阶段 1102 将与广告相关联的着陆页分割成训练着陆页和测试着陆页。例如，参考图 4，可以将着陆页和 URL 404 分成可以被用作训练示例来训练分类模型 402 的训练着陆页，以及可以被用来测试分类模型 402 的着陆页。

[0116] 阶段 1104 针对训练着陆页的侵入特征迭代训练分类模型。例如，使用由特征提取引擎 406 从自着陆页和 URL 404 获取的训练着陆页提取的特征，系统 400 可以迭代训练分类模型 402。训练可以由控制评估 408 和机器学习引擎 410 的组合来执行。

[0117] 阶段 1106 针对测试着陆页的侵入特征迭代测试分类模型直到出现测试停止事件。例如，使用由特征提取引擎 406 从自着陆页和 URL 404 获取的测试着陆页提取的特征，系统 400 可以迭代测试分类模型 402。测试可以由机器学习引擎 410 执行。在测试期间，可以诸如通过使用线性回归模型来调整特征权重和侵入特征之间的关联。可以迭代地重复阶段 1104 和 1106，例如直到出现测试停止事件，诸如确定特征权重足够好。

[0118] 阶段 1108 将特征权重和侵入特征的关联存储在分类模型中，该特征权重和侵入特征的关联取自迭代训练和测试。例如，可以将阶段 1104 和 1106 迭代生成的特征权重和侵入特征之间的关联存储在分类模型 402 中。

[0119] 在本专利文档中描述的装置、方法、流程图和结构化框图可以在包括程序代码的计算机处理系统中实现，所述程序代码包括计算机处理系统可执行的程序指令。还可以使用其它实施方式。此外，还可以利用在本专利文档中描述的流程图和结构化框图来实现对应的软件结构和算法以及它的等价物，所述流程图和结构化框图描述了特定方法和 / 或支持步骤的对应动作和支持公开的结构化装置的对应功能。

[0120] 所撰写的描述阐述了本发明的最佳模式并且提供了描述本发明的示例使得本领域技术人员能够制造和使用本发明。所撰写的描述不是将本发明限制在所阐述的精确的术语。因此，尽管已经参考上面阐述的示例详细描写了本发明，但是本领域技术人员可以在不背离本发明的范围的情况下，对他示例进行变更、改型和变化。

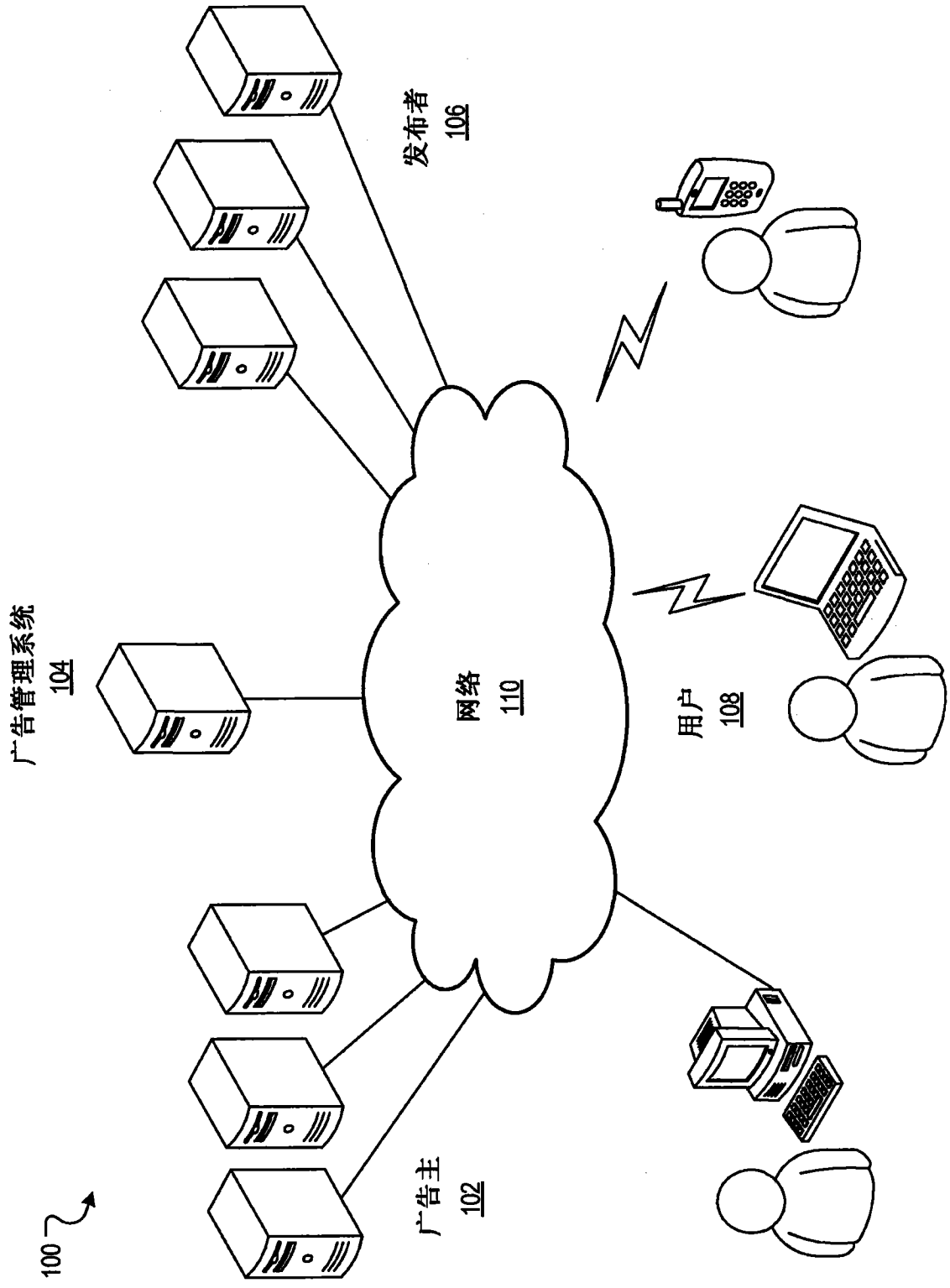


图 1

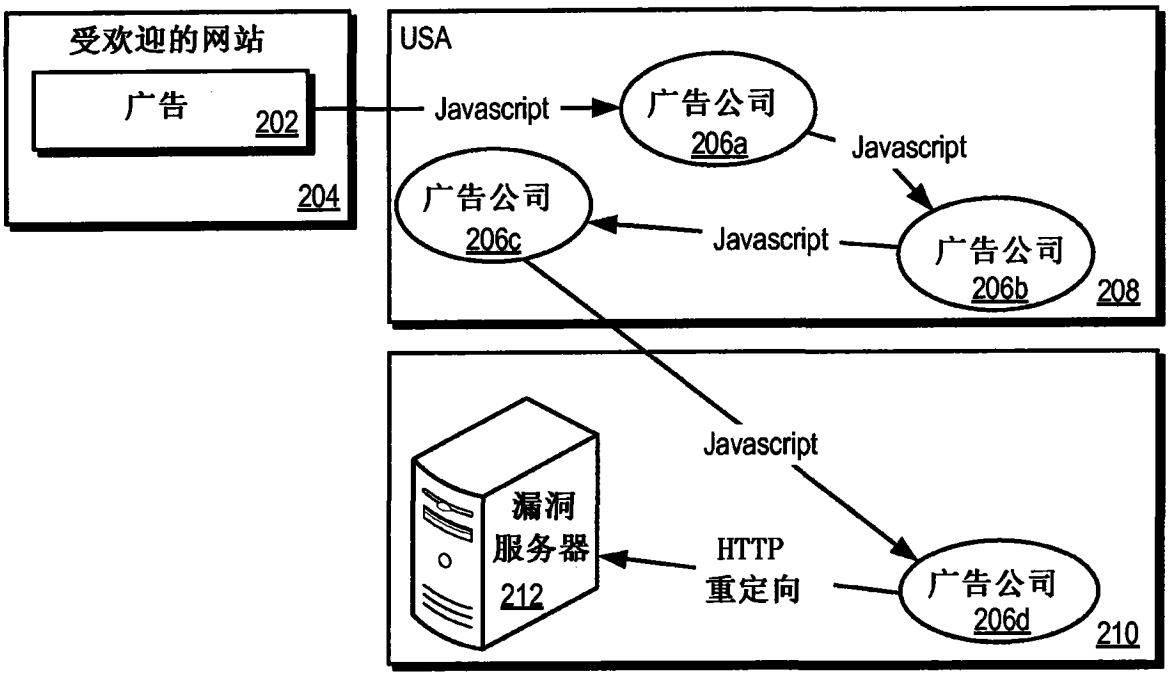

受欢迎的网站
广告
202
204
USA
Javascript
广告公司
206a
Javascript
广告公司
206c
Javascript
广告公司
206b
208
Javascript
漏洞
服务器
212
HTTP
重定向
广告公司
206d
210

图 2

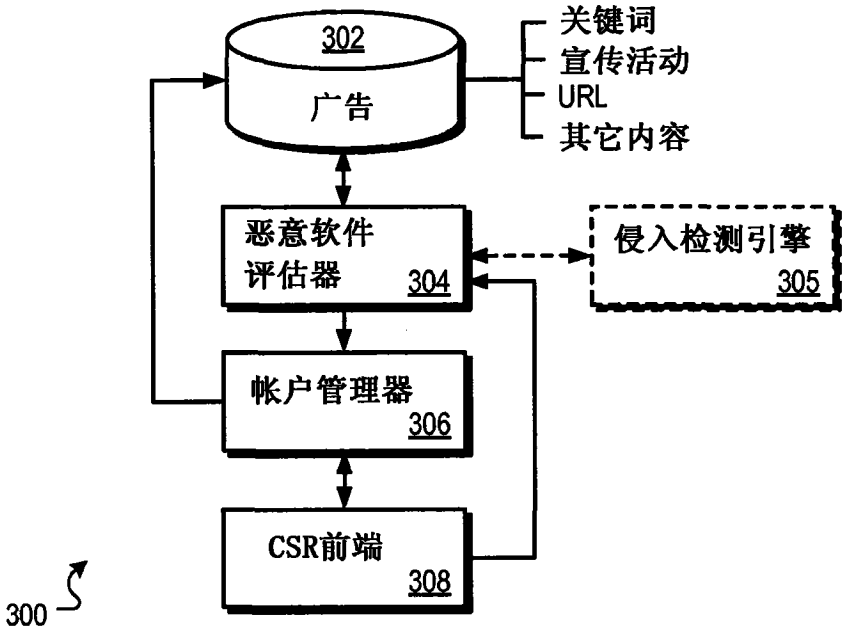

302
广告
关键词
宣传活动
URL
其它内容
恶意软件
评估器
304
侵入检测引擎
305
帐户管理器
306
CSR前端
308
300

图 3

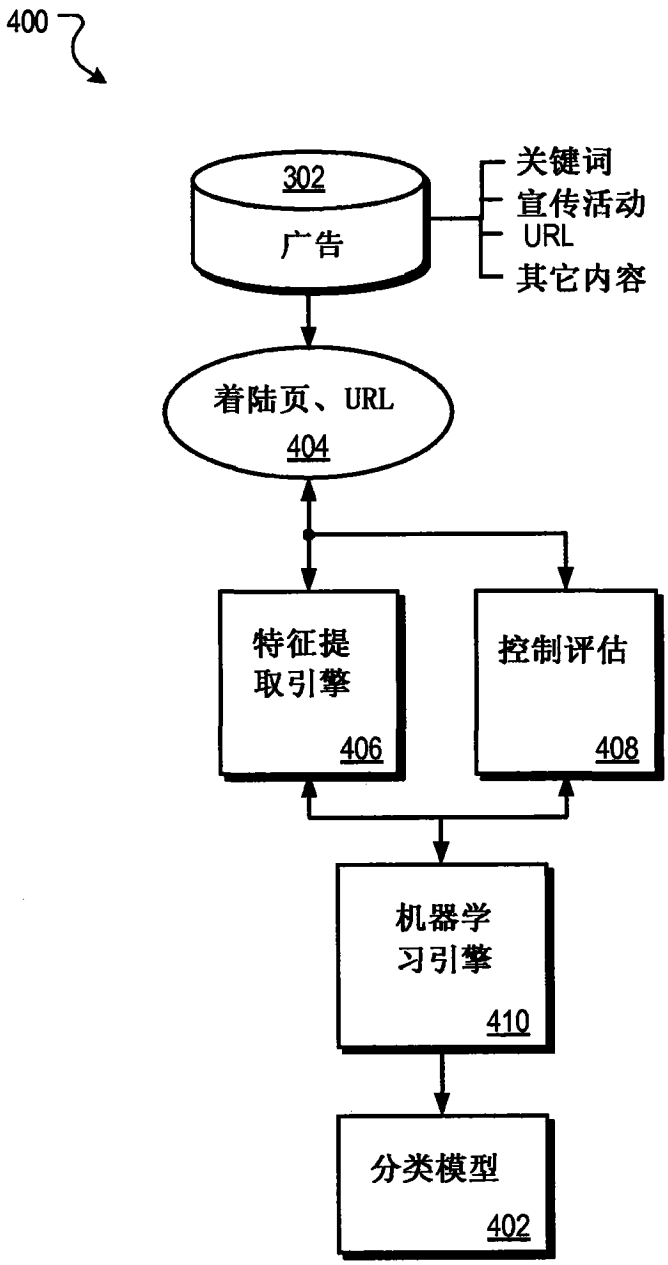

400
302
广告
关键词
宣传活动
URL
其它内容
着陆页、URL
404
特征提取引擎
406
控制评估
408
机器学习引擎
410
分类模型
402

图 4

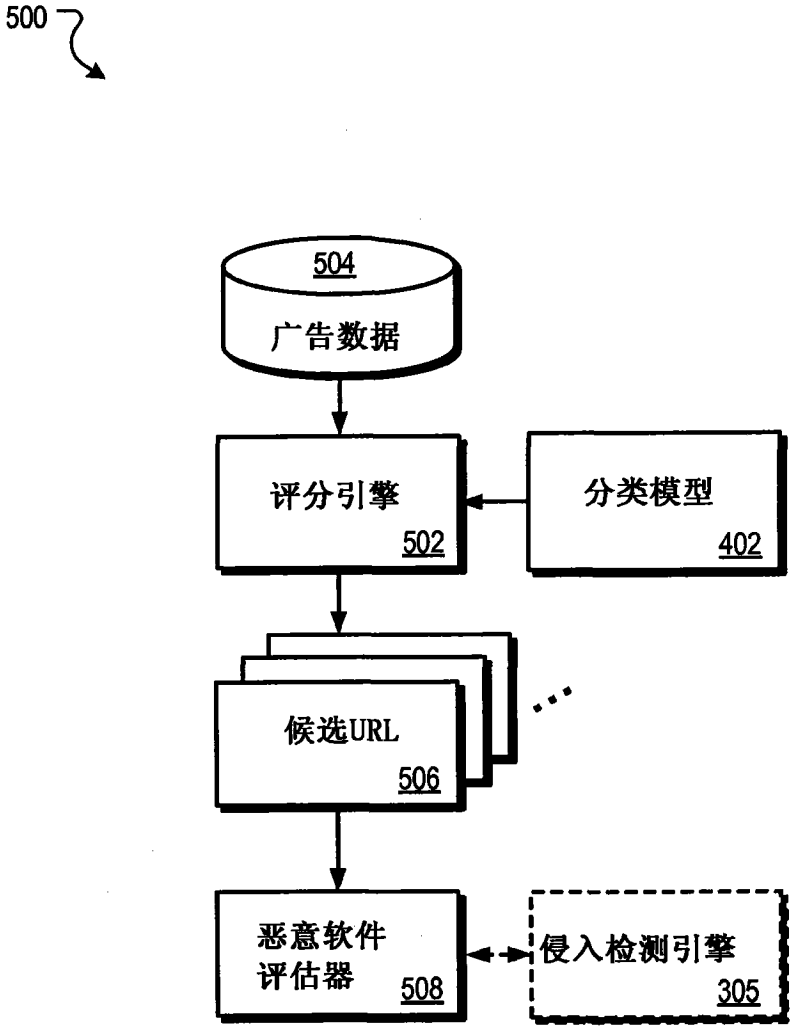

500
504
广告数据
评分引擎
502
分类模型
402
候选URL
506
恶意软件
评估器
508
侵入检测引擎
305

图 5

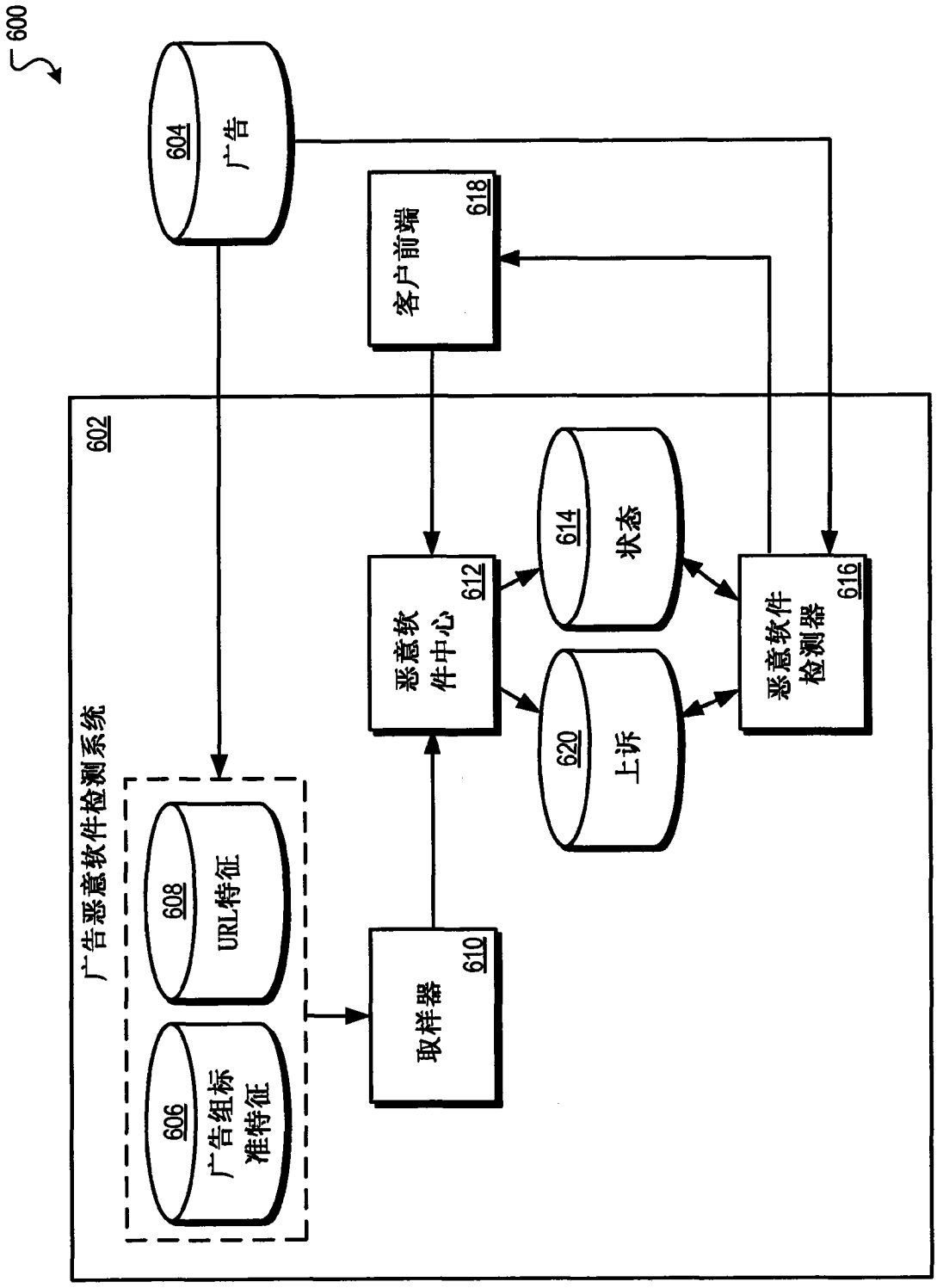

600
604
广告
618
客户前端
602
614
状态
612
恶意软件中心
616
恶意软件检测器
620
上诉
广告恶意软件检测系统
608
URL特征
610
取样器
606
广告组标准特征

图 6

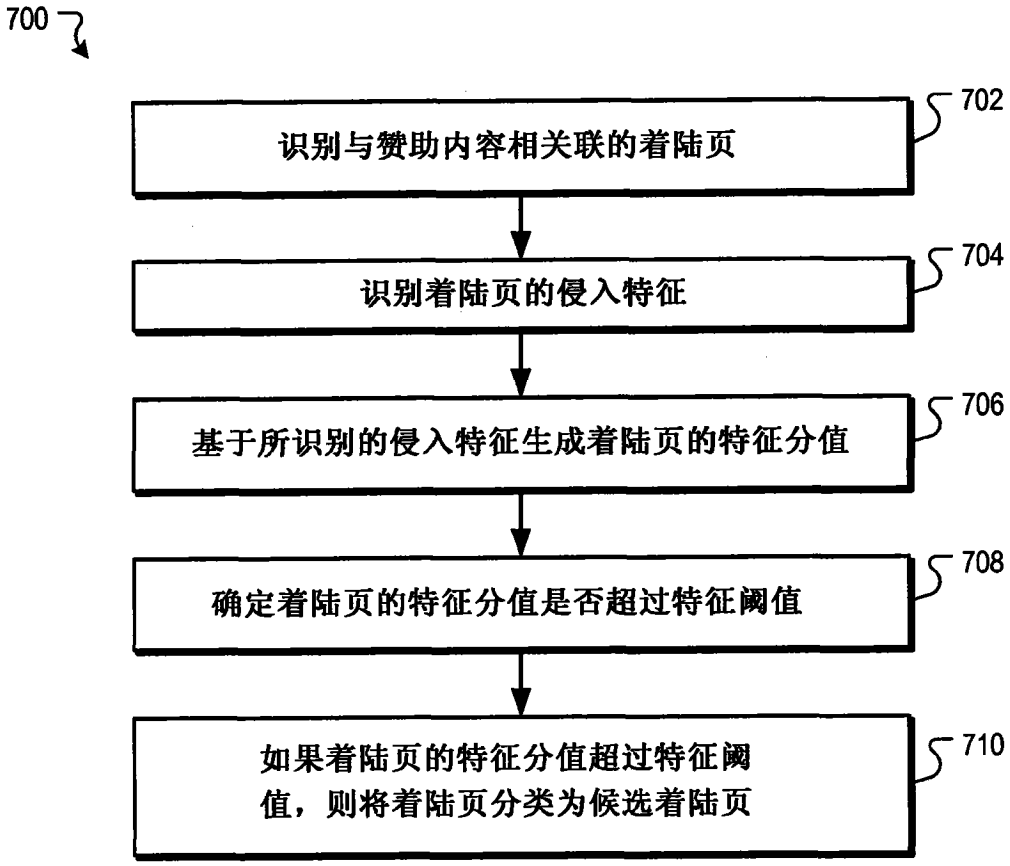

700
识别与赞助内容相关联的着陆页
702
识别着陆页的侵入特征
704
基于所识别的侵入特征生成着陆页的特征分值
706
确定着陆页的特征分值是否超过特征阈值
708
如果着陆页的特征分值超过特征阈值，则将着陆页分类为候选着陆页
710

图 7

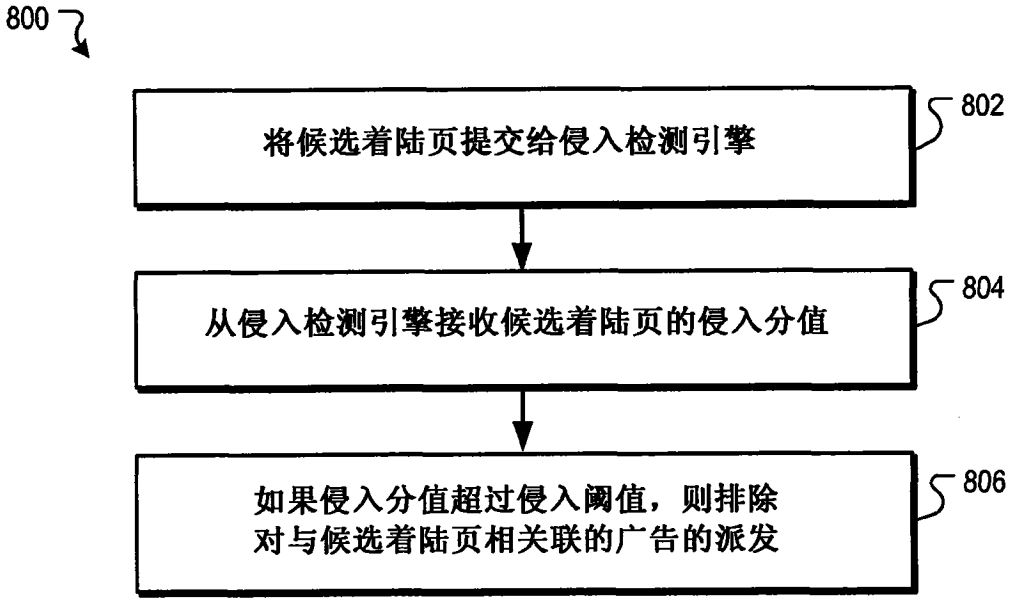

800
将候选着陆页提交给侵入检测引擎
802
从侵入检测引擎接收候选着陆页的侵入分值
804
如果侵入分值超过侵入阈值，则排除对与候选着陆页相关联的广告的派发
806

图 8

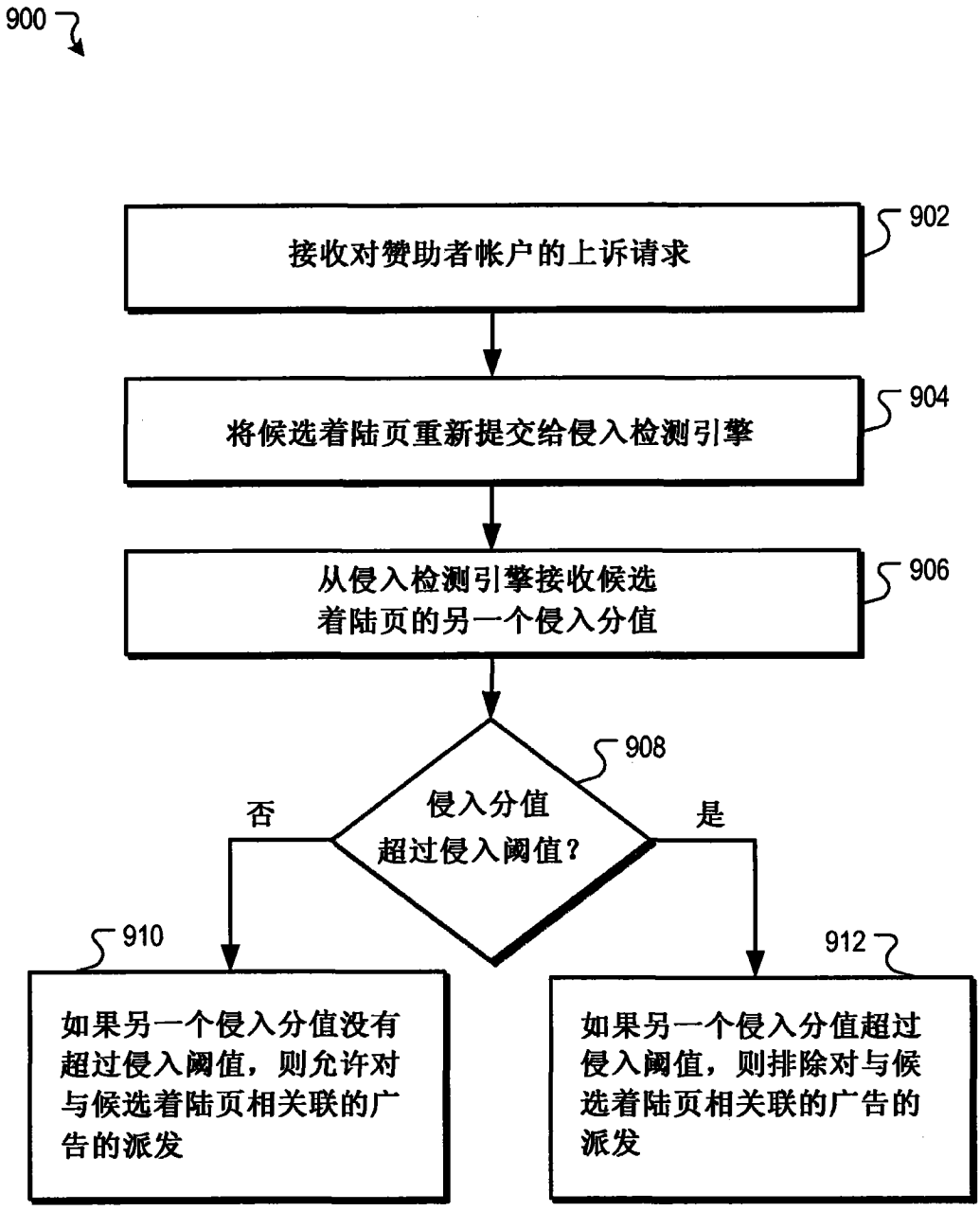

900
接收对赞助者帐户的上诉请求
902
将候选着陆页重新提交给侵入检测引擎
904
从侵入检测引擎接收候选着陆页的另一个侵入分值
906
侵入分值超过侵入阈值？
908
否
是
910
如果另一个侵入分值没有超过侵入阈值，则允许对与候选着陆页相关联的广告的派发
912
如果另一个侵入分值超过侵入阈值，则排除对与候选着陆页相关联的广告的派发

图 9

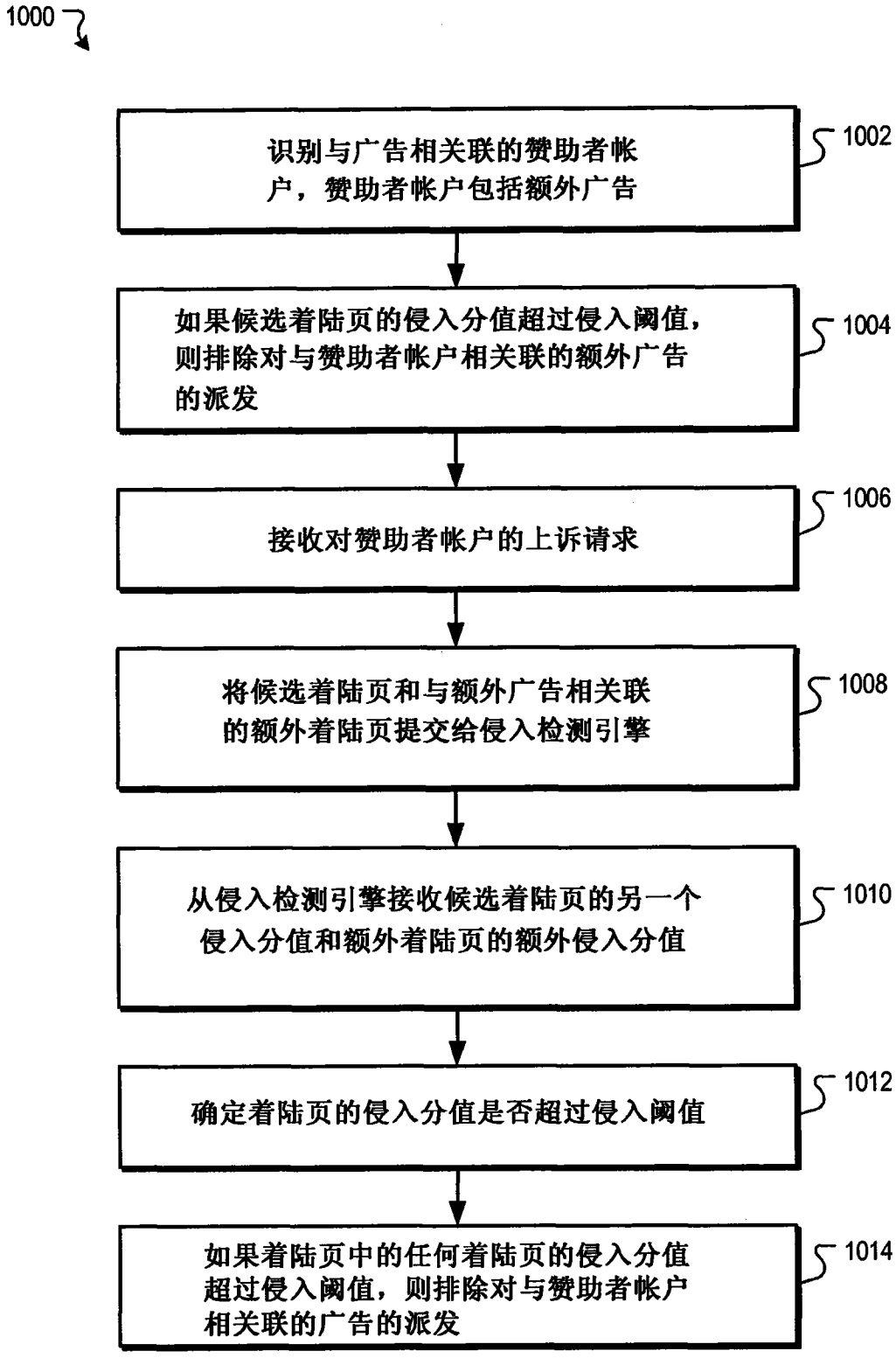

1000
识别与广告相关联的赞助者帐户，赞助者帐户包括额外广告
1002
如果候选着陆页的侵入分值超过侵入阈值，则排除对与赞助者帐户相关联的额外广告的派发
1004
接收对赞助者帐户的上诉请求
1006
将候选着陆页和与额外广告相关联的额外着陆页提交给侵入检测引擎
1008
从侵入检测引擎接收候选着陆页的另一个侵入分值和额外着陆页的额外侵入分值
1010
确定着陆页的侵入分值是否超过侵入阈值
1012
如果着陆页中的任何着陆页的侵入分值超过侵入阈值，则排除对与赞助者帐户相关联的广告的派发
1014

图 10

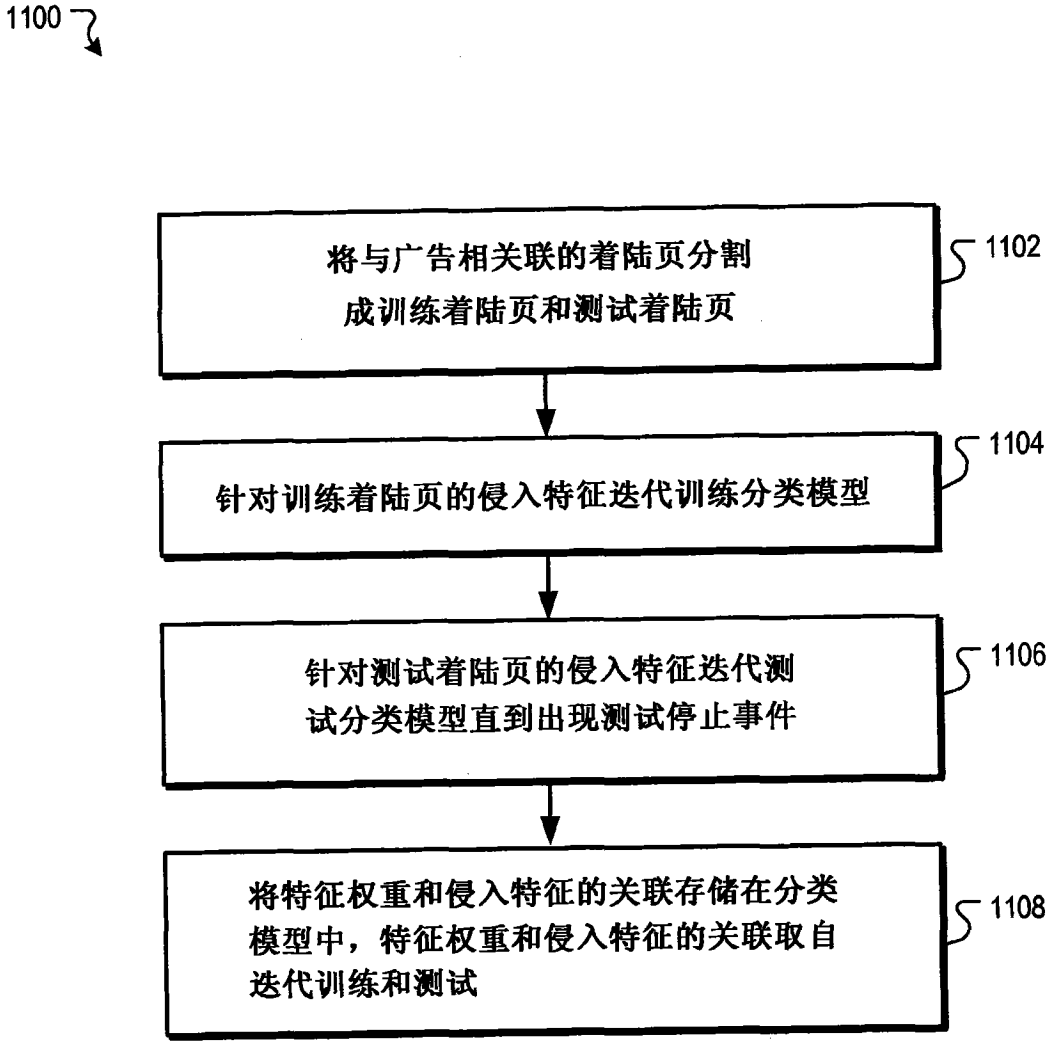

1100
将与广告相关联的着陆页分割成训练着陆页和测试着陆页
1102
针对训练着陆页的侵入特征迭代训练分类模型
1104
针对测试着陆页的侵入特征迭代测试分类模型直到出现测试停止事件
1106
将特征权重和侵入特征的关联存储在分类模型中，特征权重和侵入特征的关联取自迭代训练和测试
1108

图 11