(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 117216797 B
(45)授权公告日 2024.05.31

(21)申请号 202311243250.3

(22)申请日 2023.09.25

(65)同一申请的已公布的文献号
申请公布号 CN 117216797 A

(43)申请公布日 2023.12.12

(30)优先权数据
10202260284Y 2022.12.01 SG

(73)专利权人 丰立有限公司
地址 新加坡新加坡市

(72)发明人 陈美玲 拉贾莫汉·帕拉玛思万 陈鸿泉

(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
专利代理师 夏凯 张伟峰

(51)Int.Cl.
*G06F 21/62*(2013.01)
*G06F 21/57*(2013.01)
*G06F 21/56*(2013.01)

(56)对比文件
CN 110362995 A,2019.10.22
US 2019042743 A1,2019.02.07
CN 114896594 A,2022.08.12
CN 115001781 A,2022.09.02
KR 20150019088 A,2015.02.25
US 11010472 B1,2021.05.18

审查员 冯慧萍

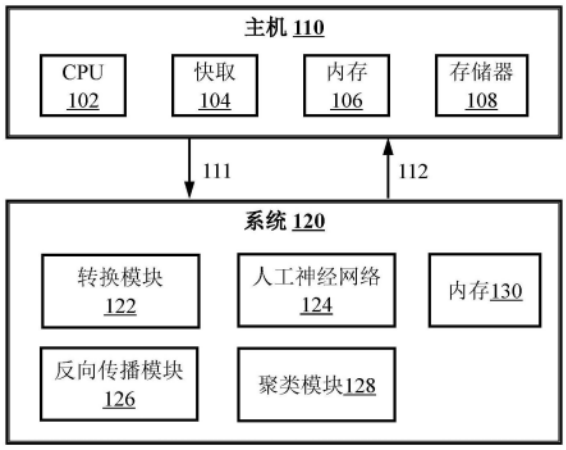
权利要求书3页 说明书11页 附图5页

(54)发明名称

# 保护数据文件的系统及方法

(57)摘要

本发明提供一种用于保护从一系列数据文件所选的数据文件的系统及方法。所述系统包括转换模块、人工神经网络(Artificial Neural Network)、聚类模块,以及反向传播模块。这些模块被设置以识别包含恶意软件或异常数据文件。当侦测到此类数据文件时,所述系统启动一系列措施以识别可能受到侦测到的恶意软件类似影响的其它数据文件。这些数据文件随即被保护,以避免恶意软件影响主机及/或与所述主机连接的任何储存设备/周边设备。




CN 117216797 B

1.一种保护数据文件的系统,其特征在于,所述系统包括:

转换模块,其包括从一系列数据文件转换的简化列阶梯形RREF矩阵组,借此所述一系列数据文件被映射至所述RREF矩阵组,其中所述转换模块被设置为:

从所述系统提取更改的数据文件并将所述更改的数据文件转换为更改的RREF矩阵组;

将所述更改的数据文件映射至所述更改的RREF矩阵组,并且将所述更改的RREF矩阵组提供至经训练的人工神经网络ANN模块;

其中,所述经训练的ANN模块被设置为:

判断所述更改的RREF矩阵组是否包括恶意活动,从而当判断出所述更改的RREF矩阵组包括恶意活动时,将所述更改的RREF矩阵组以及所述更改的数据文件移到所述系统的内存中的虚拟安全区域;

聚类模块,其被设置为:

使用反向传播模块从所述RREF矩阵组提取所述更改的RREF矩阵组的变体;

聚类所述更改的RREF矩阵组的所提取的变体以及所述更改的RREF矩阵组,以识别包含恶意活动的RREF矩阵组;

将识别出的包含恶意活动的所述RREF矩阵组提供至所述经训练的ANN模块,从而所述经训练的ANN模块被设置以识别与所识别出的所述RREF矩阵组相关的恶意活动的类型;以及

根据与数据文件相关的、通过所述经训练的ANN模块所识别的恶意活动的类型,保护被映射到识别出的、包含恶意活动的所述RREF矩阵组的所述数据文件。

2.根据权利要求1所述的系统,其特征在于,其中所述转换模块将所述更改的数据文件转换为所述更改的RREF矩阵组包括,所述转换模块被设置,以:

将所述更改的数据文件转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框;

使用第一线性函数将所述中间数据框转换为矩阵组;以及

将所述矩阵组转换为所述更改的RREF矩阵组。

3.根据权利要求1所述的系统,其特征在于,其中所述聚类模块进一步被设置,以将不包含恶意活动的RREF矩阵组提供至所述转换模块,从而所述转换模块被进一步设置为将所提供的RREF矩阵组转换为数据文件,并且将所述数据文件映射至所提供的RREF矩阵组。

4.根据权利要求1所述的系统,其特征在于,其中所述更改的RREF矩阵组的所述变体包括具有不同时间戳的相关的RREF矩阵组。

5.根据权利要求1所述的系统,其特征在于,其中所述一系列数据文件至所述RREF矩阵组的所述映射,以及所述更改的数据文件至所述更改的RREF矩阵组的所述映射,是针对每一个数据文件或每一个更改的数据文件、通过将哈希函数应用于所述数据文件的位址以及所述数据文件的语境元数据层完成,以及通过使用所述哈希函数的结果将每一个数据文件连结至每一个RREF矩阵组完成。

6.根据权利要求1所述的系统,其特征在于,其中所述经训练的ANN模块进一步被设置以进一步根据识别出的、包含恶意活动的所述RREF矩阵组,训练所述ANN模块中的所述神经网络。

7.根据权利要求1所述的系统,其特征在于,其中所述转换模块从所述一系列数据文件

转换为所述RREF矩阵组包括,所述转换模块被设置,以:

将所述数据文件中的每一个转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框;

使用第一线性函数将所述中间数据框转换为矩阵组;以及

将所述矩阵组转换为RREF矩阵组。

8.一种使用系统保护一系列数据文件的方法,其特征在于,所述系统包括转换模块,所述转换模块包括从所述一系列数据文件转换的简化列阶梯形RREF矩阵组,其中所述一系列数据文件被映射至所述RREF矩阵组,所述方法包括以下步骤:

使用所述转换模块从所述系统提取更改的数据文件,以及将所述更改的数据文件转换为更改的RREF矩阵组;

使用所述转换模块将所述更改的数据文件映射至所述更改的RREF矩阵组,以及将所述更改的RREF矩阵组提供至经训练的人工神经网络ANN模块;

使用所述经训练的ANN模块判断所述更改的RREF矩阵组是否包括恶意活动,其中当所述更改的RREF矩阵组被判断为包括恶意活动时,将所述更改的RREF矩阵组以及所述更改的数据文件移到所述系统的内存中的虚拟安全区域;

使用反向传播模块从所述RREF矩阵组提取所述更改的RREF矩阵组的变体;

使用聚类模块聚类所述更改的RREF矩阵组的所提取的变体以及所述更改的RREF矩阵组,以识别包含恶意活动的RREF矩阵组;

使用所述聚类模块将识别出的、包含恶意活动的所述RREF矩阵组提供至所述经训练的ANN模块,其中所述经训练的ANN模块被设置以识别与所识别出的所述RREF矩阵组相关的恶意活动的类型;以及

根据与数据文件相关的、通过所述经训练的ANN模块所识别的恶意活动的类型,使用所述聚类模块保护被映射到识别出的、包含恶意活动的所述RREF矩阵组的所述数据文件。

9.根据权利要求8所述的方法,其特征在于,其中将所述更改的数据文件转换为所述更改的RREF矩阵组包括:

使用所述转换模块将所述更改的数据文件转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框;

使用所述转换模块,以使用第一线性函数,将所述中间数据框转换为矩阵组;以及

使用所述转换模块,将所述矩阵组转换为所述更改的RREF矩阵组。

10.根据权利要求8的方法,其特征在于,所述方法进一步包括以下步骤:

使用所述聚类模块,将不包含恶意活动的RREF矩阵组提供至所述转换模块,从而所述转换模块进一步被设置为将所提供的RREF矩阵组转换为数据文件,并且将所述数据文件映射至所提供的RREF矩阵组。

11.根据权利要求8所述的方法,其特征在于,其中所述更改的RREF矩阵组的所述变体包括具有不同时间戳的相关的RREF矩阵组。

12.根据权利要求8所述的方法,其特征在于,其中所述一系列数据文件至所述RREF矩阵组的所述映射,以及所述更改的数据文件至所述更改的RREF矩阵组的所述映射,是针对于每一个数据文件或每一个更改的数据文件、通过将哈希函数应用于所述数据文件的位址以及应用于所述数据文件的语境元数据层完成,并且使用所述哈希函数的结果将每一个数

据文件连结至每一个RREF矩阵组完成。

13.根据权利要求8所述的方法,其特征在于,其中所述方法进一步包括,根据识别出的包含恶意活动的所述RREF矩阵组,训练在所述ANN模块中的所述神经网络的步骤。

14.根据权利要求8所述的方法,其特征在于,其中所述一系列数据文件通过所述转换模块转换为所述RREF矩阵组包括以下步骤:

使用所述转换模块,将所述数据文件的每一个转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框;

使用所述转换模块,使用第一线性函数将所述中间数据框转换为矩阵组;以及

使用所述转换模块,将所述矩阵组转换为RREF矩阵组。

# 保护数据文件的系统及方法

## 技术领域

[0001] 本发明涉及一种用于保护从一系列的数据文件中所选的数据文件的系统及方法。

## 背景技术

[0002] 恶意软件是在现代计算领域中的持续威胁。由于越来越多计算设备通过无线或有线网络连接,恶意软件可以轻易地通过网络传播,其感染所有连线的计算设备。受感染的电脑会降低个人和组织的生产力,并且还可能造成无法挽回的损失。例如,受感染的电脑可能会有被破坏的操作系统或损毁的数据,从而使其运行速度变慢或使其完全无法运行。在最糟的情况下,有价值的数据及信息可能被盗用、劫持、滥用或被用于勒索。此外,此类受感染的系统可能会继续将恶意软件传播到网络之外的其它设备。

[0003] 已知的防恶意软件可能会尝试通过侦测这些已知恶意软件的签名或找出与其它已知恶意软件相关的程序中的模式来识别和侦测已知恶意软件。此种反恶意软件方法根据基于签名的侦测系统,将执行方式及数据与已知恶意软件的识别特征相比较而运作。然而,零日攻击通常不符合任何已知恶意软件的识别特征。因此,现有的防恶意软件解决方案可能无法侦测零日攻击,这与已知恶意软件不同,并且该不同不可预测。

[0004] 如同上文所提及的,有价值的数据及信息可能会被恶意软件盗用,并且此类数据通常是无可取代的。因此,通常会采取预防措施,例如经常备份数据、或对数据进行加密并将加密的数据储存在安全的环境中。

[0005] 然而,此类措施仍面临许多问题,因为备份的数据可能被非法入侵或损毁、可能出现硬件故障、意外损坏、勒索软件或恶意软件感染、或被意外删除。在勒索软件的情形中,攻击者将会加密用户的文件并向受害者勒索赎金以恢复对数据的存取。此类问题中最常见的情况之一是网络钓鱼垃圾邮件。在网络钓鱼垃圾邮件中,毫无戒心的用户下载并执行了损坏的附件。一旦这些文件被执行,它们就会通过自行安装相关脚本以及破坏内含在其中的数据来控制受攻击设备的操作系统。

[0006] 某些形式的恶意脚本不需要通过用户就能执行。此种脚本可能会使用操作系统中的安全漏洞来感染内含在其中的文件及数据,从而对计算设备造成损害。此外,还有称为泄漏软件以及隐私挟持软件的网络攻击。在此类攻击中,获得计算设备存取权限的恶意行为者会威胁将内含在所述计算设备储存介质中的敏感数据公诸于世。

[0007] 恶意第三方还可能使用特洛伊木马程序攻击来破坏、损毁和窃取用户的数据。特洛伊木马程序还会产生病毒、蠕虫、间谍软件、僵尸、僵尸网络、逻辑炸弹和后门程序。一般而言,所有这些不同类型的恶意网络攻击的唯一目的都是损害用户数据的安全性。这些攻击通常会破坏操作系统、操作系统的核心层以及储存数据的位址指针,从而破坏所述数据本身。

[0008] 基于以上原因,本领域技术人员不断努力提出一种能够通过识别和侦测可能包含恶意软件的数据文件来保护计算设备中的档案的系统和方法。

## 发明内容

[0009] 根据本发明的实施例提供的系统和方法解决了上述及其它问题,并且在此项技术取得了进步。

[0010] 根据本发明的系统及方法的实施例的第一优点是,所述系统被设置为保护所有类型的文件,而不论所述文件是可执行文件或非可执行文件。

[0011] 根据本发明的系统及方法的实施例的第二优点是,所述系统被设置为,将数据文件存为非授权用户难以操控和重建的格式,即使主机计算设备的操作系统已经受损。

[0012] 根据本发明的系统及方法的实施例的第三优点是,一旦文件根据本发明的实施例而被储存,所述系统能够通过被用来将所述RREF矩阵映射至所述被储存的文件的映射函数,轻易且有效率地恢复损毁的文件。

[0013] 根据本发明的系统及方法的实施例的第四优点是,所述系统被设置为在使用经训练的神经网络分析文件之前,在所述文件被转换为数值表达时,侦测内含在所述数据文件内的恶意软件,以及零日漏洞类型的恶意软件。

[0014] 根据本发明的系统及方法的实施例的第五优点是,所述系统被设置为将被判断为包含恶意软件的文件储存在所述内存中的虚拟安全区域内,借此所述被储存的文件只能够被经授权的用户提取。

[0015] 以上的优点由根据本发明用以下方式操作的系统及/或方法的实施例所提供。

[0016] 根据第一方面,本发明提供一种用于保护一系列数据文件的系统,该系统包括转换模块及聚类模块。所述转换模块包括从所述一系列数据文件转换的简化列阶梯形(Reduced Row Echelon Form,RREF)矩阵组,借此所述一系列数据文件被映射至所述RREF矩阵组。所述转换模块被设置为:从所述系统提取更改的数据文件并且将所述更改的数据文件转换为更改的RREF矩阵组、将所述更改的数据文件映射至所述更改的RREF矩阵组,以及将所述更改的RREF矩阵组提供至经训练的人工神经网络(ANN)模块。所述经训练的ANN模块被设置为:判断所述更改的RREF矩阵组是否包括恶意活动。当判断出所述更改的RREF矩阵组包括恶意活动时,将所述更改的RREF矩阵组以及所述更改的数据文件移到所述系统的内存中的虚拟安全区域。所述聚类模块被设置为:使用反向传播模块从所述RREF矩阵组提取所述RREF矩阵组的变体、聚类所述更改的RREF矩阵组的所提取的变体以及所述更改的RREF矩阵组以识别包含恶意活动的RREF矩阵组、将识别出的包含恶意活动的RREF矩阵组提供至所述经训练的ANN模块,从而所述经训练的ANN模块被设置以识别与所述识别出的RREF矩阵组相关的恶意活动的类型,以及根据与数据文件相关的、通过所述经训练的ANN模块所识别的恶意活动的所述类型,保护被映射到所述识别出的、包含恶意活动的RREF矩阵组的所述数据文件。

[0017] 根据本发明第一方面的实施例,所述更改的数据文件通过所述转换模块的所述转换为所述更改的RREF矩阵组包括所述转换模块被设置,以将所述更改的数据文件转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框、使用第一线性函数将所述中间数据框转换为矩阵组,以及将所述矩阵组转换为所述更改的RREF矩阵组。

[0018] 根据本发明第一方面的实施例,所述聚类模块进一步被设置,以将不包含恶意活动的RREF矩阵组提供至所述转换模块,从而所述转换模块被进一步设置为将所提供的RREF矩阵组转换为数据文件,并且将所述数据文件映射至所提供的RREF矩阵组。

[0019] 根据本发明第一方面的实施例,所述更改的RREF矩阵组的变体包括具有不同时间戳的相关的RREF矩阵组。

[0020] 根据本发明第一方面的实施例,所述一系列的数据文件至所述RREF矩阵组的所述映射、以及所述更改的数据文件至所述更改的RREF矩阵组的所述映射,是针对于每一个数据文件或每一个更改的数据文件完成的。其中所述映射通过将哈希函数应用于所述数据文件的位址以及所述数据文件的语境元数据层完成,以及通过使用此哈希函数的结果将每一个数据文件连结到每一个RREF矩阵组完成。

[0021] 根据本发明的第一方面的实施例,所述经训练的ANN模块进一步被设置,以进一步根据所述识别出的、包含恶意活动的RREF矩阵组,训练在所述ANN模块中的所述神经网络。

[0022] 根据本发明的第一方面的实施例,所述一系列的数据文件通过所述转换模块的所述转换为所述RREF矩阵组包括,所述转换模块被设置,以将所述数据文件的每一个转换为一中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框、使用第一线性函数将所述中间数据框转换为一矩阵组,以及将所述矩阵组转换为RREF矩阵组。

[0023] 根据第二方面,本发明提供使用系统的、用于保护一系列数据文件的方法。该系统包括转换模块,所述转换模块包括从所述一系列的数据文件转换的简化列阶梯形(RREF)矩阵组,其中所述一系列数据文件被映射至所述RREF矩阵组,所述方法包括以下步骤:(1)使用所述转换模块来从所述系统提取更改的数据文件,以将所述更改的数据文件转换为更改的RREF矩阵组;(2)使用所述转换模块将所述更改的数据文件映射至所述更改的RREF矩阵组,以及提供所述更改的RREF矩阵组至经训练的人工神经网络(ANN)模块;(3)使用所述经训练的ANN模块判断所述更改的RREF矩阵组是否包括恶意活动,当判断出所述更改的RREF矩阵组包括恶意活动时,将所述更改的RREF矩阵组以及所述更改的数据文件移到所述系统的内存中的虚拟安全区域;(4)使用反向传播模块,从所述RREF矩阵组提取更改的所述RREF矩阵组的变体;(5)使用聚类模块聚类所述更改的RREF矩阵组的所提取的变体以及所述更改的RREF矩阵组,以识别包含恶意活动的RREF矩阵组;(6)使用所述聚类模块将所述识别出的包含恶意活动的RREF矩阵组提供至所述经训练的ANN模块,其中所述经训练的ANN模块被设置以识别与所述识别出的RREF矩阵组相关的恶意活动的类型,以及(7)根据与数据文件相关的、通过所述经训练的ANN模块所识别的恶意的活动的所述类型,并且使用所述聚类模块保护被映射到所述识别出的包含恶意的活动的RREF矩阵组的所述数据文件。

[0024] 根据本发明的第二方面的实施例,所述将更改的数据文件转换为所述更改的RREF矩阵组的转换包括:使用所述转换模块将所述更改的数据文件转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框、使用所述转换模块以使用第一线性函数将所述中间数据框转换为矩阵组,以及使用所述转换模块将所述矩阵组转换为所述更改的RREF矩阵组。

[0025] 根据本发明第二方面的实施例,所述方法进一步包括以下步骤:使用所述聚类模块将不包含恶意活动的RREF矩阵组提供至所述转换模块,从而所述转换模块进一步被设置为将所提供的RREF矩阵组转换为数据文件、以及将所述数据文件映射至所提供的RREF矩阵组。

[0026] 根据本发明第二方面的实施例,所述更改的RREF矩阵组的所述变体包括具有不同时间戳的相关的RREF矩阵组。

[0027] 根据本发明第二方面的实施例,所述一系列数据文件至所述RREF矩阵组的所述映射、以及所述更改的数据文件至所述更改的RREF矩阵组的所述映射,是针对于每一个数据文件或每一个更改的数据文件完成的,其中所述映射通过将哈希函数应用于所述数据文件的位址以及应用于所述数据文件的语境元数据层完成,并且使用此哈希函数的结果将每一个数据文件连结至每一个RREF矩阵组完成。

[0028] 根据本发明的第二方面的实施例,所述方法进一步包括根据所述识别出的包含恶意活动的RREF矩阵组训练在所述ANN模块中的所述神经网络的步骤。

[0029] 根据本发明第二方面的实施例,所述一系列的数据文件通过所述转换模块的所述转换为所述RREF矩阵组包括以下的步骤:使用所述转换模块将所述数据文件的每一个转换为中间数据框,其中所述中间数据框包括多媒体数据框或基于字符的数据框、使用所述转换模块以使用第一线性函数来将所述中间数据框转换为矩阵组,以及使用所述转换模块将所述矩阵组转换为RREF矩阵组。

## 附图说明

[0030] 以上及其它的问题可以通过根据在本文详细说明中所描述并且在参照以下附图中所展示的、本发明的系统及方法的方案及优点得到解决。

[0031] 图1是根据本发明实施例的一种用于保护主机数据文件的系统的框图;

[0032] 图2是根据本发明实施例的代表被设置在模块或计算设备内用于执行实施例的部件的系统的框图;

[0033] 图3是根据本发明实施例的用于将一系列数据文件转换为简化列阶梯形(RREF)矩阵组的流程图;

[0034] 图4是根据本发明实施例的用于保护数据文件的流程图;

[0035] 图5是根据本发明的实施例的数据文件转换为中间数据框的示意图;以及

[0036] 图6是根据本发明的实施例的中间文件转换为简化列阶梯矩阵的示意图。

## 具体实施方式

[0037] 本发明涉及用于保护从一系列数据文件所选的数据文件的系统及方法。所述系统包括转换模块、人工神经网络(Artificial Neural Network ANN)、聚类模块、以及反向传播模块。这些模块被设置以识别包含恶意软件或异常信息的数据文件。当此类数据文件被侦测到时,所述系统将会启动一系列的相应措施,以识别可能类似受到所述侦测到的恶意软件影响的其它数据文件。这些数据文件随即被保护,以避免所述恶意软件影响主机及/或任何连结至所述主机的储存设备/周边设备。

[0038] 具体地,转换模块包括已从所述一系列数据文件转换的简化列阶梯形(Reduced Row Echelon Form RREF)矩阵组,借此将所述一系列数据文件映射至所述RREF矩阵组。转换模块被设置为从所述系统提取更改的数据文件,并且将所述更改的数据文件转换为更改的RREF矩阵组。所述更改的数据文件继而被映射至所述更改的RREF矩阵组,并且随即被提供至经训练的人工神经网络(ANN)模块。在接收到所述更改的RREF矩阵组之后,所述经训练的ANN模块随即被设置为判断所述更改的RREF矩阵组是否包括恶意活动,从而当判断出所述更改的RREF矩阵组包括恶意活动时,将所述更改的RREF矩阵组以及所述更改的数据文件

移到所述系统的内存中的虚拟安全区域。聚类模块随后使用反向传播模块从所述RREF矩阵组提取所述更改的RREF矩阵组的变体,并且继续将所述更改的RREF矩阵组的所提取的变体以及所述更改的RREF矩阵组聚类在所述安全区域中,以识别包含恶意活动的RREF矩阵组。一旦被识别后,所述识别出的包含恶意活动的RREF矩阵组随之被提供至所述经训练的ANN模块,并且被映射到所述识别出的包含恶意活动的RREF矩阵组的数据文件将会被保护。

[0039] 现在将参考如附图对所呈现的本发明实施例作详细描述。以下的说明阐述多个特定的特征,以便于提供本发明的实施例的完整理解。然而,对于本领域技术人员而言,实施例可以在不包括某些或全部特征的方案中实现。此种实施例也应被理解为落入本发明的范围内。简明起见,某些已知的程序步骤及/或结构在下文中可能未做详述。本领域技术人员可以参考对应的引用文,以免不必要地模糊理解本发明。

[0040] 此外,本领域技术人员将会认识到,本文所述的某个或某些功能单元在说明书全文中以模块形式做示例性说明。本领域技术人员亦将会认识到,模块可以是电路、逻辑晶片或任意类型的离散的部件的形式实施。进一步地,本领域技术人员还将认识到,模块可以通过软件实施,从而可以通过各种处理器架构执行。在本发明的实施例中,模块还可包括电脑指令、硬件或可执行的代码,其可以根据接收到的指令指示电脑处理器实行一系列的事件。所述模块的实施选择可以是本领域技术人员的设计选择,因而并不以任何方式限制本发明的范围。

[0041] 根据本发明实施例的用于保护数据文件的示例的程序或方法可以由以下步骤阐述。

[0042] 步骤1:使用转换模块从所述系统提取更改的数据文件,并且将所述更改的数据文件转换为更改的RREF矩阵组;将所述更改的数据文件映射至所述更改的RREF矩阵组,并且将所述更改的RREF矩阵组提供至经训练的人工神经网络(ANN)模块,借此,所述转换模块包括从所述系列的数据文件转换的简化列阶梯形(RREF)矩阵组,因而借此所述一系列数据文件被映射至所述RREF矩阵组,

[0043] 步骤2:使用所述经训练的ANN模块判断所述更改的RREF矩阵组是否包括恶意活动。当判断出的是所述更改的RREF矩阵组包括恶意活动时,将所述更改的RREF矩阵组以及所述更改的数据文件移到所述系统内存中的虚拟安全区域,

[0044] 步骤3:使用反向传播模块,从所述RREF矩阵组提取所述更改的RREF矩阵组的变体,

[0045] 步骤4:使用聚类模块,聚类所述更改的RREF矩阵组的所提取的变体以及所述更改的RREF矩阵组,以识别包含恶意的活动的RREF矩阵组,以及

[0046] 步骤5:使用所述聚类模块,将所述识别出的包含恶意活动的RREF矩阵组提供至所述经训练的ANN模块,并且保护被映射到所述识别出的包含恶意活动的RREF矩阵组的数据文件。

[0047] 根据本发明的实施例,上文所阐述的步骤可以通过通讯地连线于主机的系统或硬件模块(包括子模块)来实行或执行。以上的步骤亦可以通过被设置在所述主机及/或在连线的周边装置的软件模块来实行或执行。

[0048] 根据本发明实施例的一种用于保护在主机中的数据文件的系统的方块图被描绘在图1中。

[0049] 如图1所示,主机110基本可包括具有中央处理器(CPU)102、快取104、内存106及/或储存器108的任意硬件装置。主机的某些例子包括但不限于电脑、个人电子装置、精简型客户端,以及多功能设备。具体地,几乎任何种类的电脑,包含集中式大型主机、服务器或桌上型个人电脑(PC)都可被设置为主机。

[0050] CPU 102可包括任何可以处理此种指令的设备或部件,并且可包括:微处理器、微控制器(MCU)、可程序化逻辑装置或其它计算设备。换言之,CPU 102可以是由任何适当的逻辑电路所提供,其用于接收输入、根据储存在内存106中的指令来处理以及产生输出,并且其可包括具有内存可定址空间的单核心或多核心处理器。此外,内存106可包括易失性及非易失性存储装置,并且储存器108可包括(但不限于)固态驱动器(SSD)、硬盘驱动器(HDD)、光盘机或磁盘机。

[0051] 主机110通过I/O集线器(未显示)可通讯地连接于系统120。I/O集线器可包括(但不限于)任何类型的微晶片,其可被用来管理主机110内的CPU 102与各电子部件之间的数据通讯,并且管理在主机110与系统120之间交换的数据。

[0052] 系统120包括转换模块122、人工神经网络(ANN)模块124、内存130、反向传播模块126,以及聚类模块128。转换模块122设置为将数据文件转换为简化列阶梯形(RREF)矩阵,并且将RREF矩阵重建回到数据文件。人工神经网络(ANN)模块124包括经训练的神经网络。内存130包括安全及不安全的易失性及非易失性存储装置。反向传播模块126设置为识别数据文件及/或RREF矩阵组的变体。聚类模块128设置为将RREF矩阵聚类以生成RREF聚类矩阵,以识别可能包含恶意活动的RREF聚类矩阵。本领域技术人员将会认识到,系统120可包括其它类型的额外的模块及/或部件,而不脱离本发明。

[0053] 根据本发明的实施例,在最初的设置阶段期间,系统120将会从主机110获得数据文件111的整个系列、整个系列的某个部分、或所选的部分。数据文件111可包括(但不限于)任意类型的可被储存于内存106或储存器108的档案、或任何可以在主机110的模块之间内部传递或通过主机110的模块执行的档案。校验和(checksum)接着将会从具有校验和的数据文件被提取出,并且此信息随后被储存于系统120的内存130中,以供后续使用。

[0054] 在整个系列的数据文件111从主机110获得的前提下,系统120接着继续将各种许可及用户设定绑定至数据文件111的每一个。与所述数据文件的每一个相关的所述元数据(metadata)可包含有关于所述绑定的文件的真实性的信息。此外,该元数据可作为认证手段,例如诸如.PEM文件的认证凭证,以用于内部的验证程序。

[0055] 转换模块122接着继续转换所述数据文件(其可包括绑定有许可及用户设定的数据文件、或可包括未绑定的数据文件)的每一个。每个数据文件首先被转换为中间数据框,因而其随后可以轻易地被转换为矩阵组。在本发明的实施例中,中间数据框可包括多媒体数据框,其包括具有图像文件格式的框、或可包括基于字符的数据框。鉴于多媒体数据框的红-绿-蓝(RGB)本质,将它们转换为数值变得可行。在本发明的其它实施例中,所述中间数据框可替代地可包括基于字符的数据框。在本发明的进一步实施例中,中间数据框可包括所述多媒体数据框以及基于字符的数据框两者。

[0056] 不论所述中间数据框是多媒体数据框或是基于字符的数据框,其随后都被转换为矩阵组。此被描绘在图5中,其示出目标文件/数据文件502通过转换程序506被转换为矩阵快照504。图6描绘所述矩阵快照604至数值矩阵602的转换。在本发明的实施例中,此转换可

以使用预设的查找表完成,借此在所述中间数据框中的每一个字母与数值的符号及/或字元及/或像素及/或颜色都可以通过唯一的数字及/或符号表示。预设的矩阵产生演算法或函数随即可以与所述查找表一起使用,以将中间数据框转换为矩阵组。在本发明的其它实施例中,更复杂的数学函数可被用来将中间数据框转换为数值的数据框,并且这些数学函数可包括例如随机排序、旋转、顺序改变等类函数。

[0057] 在本发明的进一步实施例中,偏微分方程函数可被用于所述数值的查找表,以对所述数据做微分处理,并且使用相依的参数加密所述数据。所述相依的参数可包括哈希或加密密钥,其中的数据被参数化及分段化。

[0058] 所述中间数据框的每一个被转换为矩阵组后,转换模块122接着进一步将所述矩阵组(其可以通过线性方程表示)转换为简化列阶梯形(RREF)矩阵组,并且此转化处理可以(但不限于)通过各种的数学运算被完成。将所述矩阵组转换为RREF矩阵组的转换步骤是增加额外的加密层至所述数据,因为所述转换程序牵涉到线性方程的使用,而所述线性方程是仅有系统120才能够识别。因此,这些线性方程作用为用于系统120的独特的加密演算法。

[0059] 所述一系列数据文件111通过转换模块122被转换为其对应的RREF矩阵组后,每一个数据文件接着被映射到其RREF矩阵组,并且此映射处理可以使用(但不限于)密钥值配对函数或密钥值哈希函数完成。被用来产生所述配对的所述函数,以及所述产生的映射,与所述RREF矩阵组接着被安全地储存在内存130中。

[0060] 在本发明的另一实施例中,哈希函数可被应用于数据文件的元数据及/或所述数据文件的位址。来自所述哈希函数的结果随即可被用于将所述数据文件映射至其对应的RREF矩阵组。在所述哈希函数中,数据文件的语境元数据层的使用会增加独特的性质至所述映射函数,因为不仅所述数据文件的位址被验证,而且所参照的元数据也被验证。

[0061] 在本发明的实施例中,哈希函数可包括自然语言处理(NLP)算法以及密钥值绑定函数的组合。所述NLP演算法被设置为处理所述语境元数据,并且半标记将它们参数化以及将它们储存在特定的密钥位址的信息。上述处理改善(所述映射函数的)搜寻演算法,并且设定复杂度值(complexity value)为0(1),因为找出所述数据框的工作是包括所述系统搜寻可以与针对于所述数据文件的校验和的验证一起产生的适当的密钥。因此,此密钥在所述系统之中可以是唯一定义的。

[0062] 在传统的操作系统中,数据文件单纯被储存在所述主机的十六进制内存位址,因而指针被用来识别所述数据文件的位置。若这些指针被恶意软件损毁,损毁的指针将会使得数据文件变得无用,因为这些数据文件将不能够被所述主机提取。

[0063] 在本发明的实施例中,哈希函数的应用步骤完成后,系统120接着持续地监视主机110是否有所述一系列的数据文件111的改变。当系统120在主机110中侦测到更改的数据文件时,系统120继续从主机110提取所述更改的数据文件。在本发明的实施例中,系统120首先验证所述更改的数据文件的校验和,以判断所述更改的数据文件是否被未经授权的用户更改。若所述更改的数据文件包含损毁的校验和或不匹配的校验和,则此更改的数据文件接着被移到安全区域中,以进一步加以处理。然而,恶意的第三方此后可能会发现更改数据文件同时保持所述校验和不变的方式。因此,需要将所述更改的数据文件通过系统120作进一步处理。

[0064] 使用转换模块122,所述被提取的更改的数据文件接着被转换为更改的RREF矩阵

组。在本发明的实施例中,转换为更改的RREF矩阵组的步骤可以通过在所述更改的数据文件如前文所述的、被转换为矩阵组之前,先将所述更改的数据文件转换为中间数据框来完成。所述矩阵组接着使用数学运算被转换为其对应的更改的RREF矩阵组。转换模块122接着将所述更改的数据文件映射至所述更改的RREF矩阵组,并且将此映射安全地储存在内存130中。所述更改的RREF矩阵组接着被提供至经训练的人工神经网络(ANN)模块124。

[0065] 在本发明的其它实施例中,当数据文件被修改或更改时,被用来将所述更改的数据文件转换为更改的RREF矩阵组的、多项式方程的根,可被用于识别在所述更改的RREF矩阵组中被更改及/或改变的数字。若在所述更改的RREF矩阵组中的数字包括所述多项式方程的根的倍数,则意味着所述数据文件尚未损毁。相反地、如果所述数据文件已经被损毁,若在所述更改的RREF矩阵组中的数字不包括所述多项式方程式的根的倍数,则意味着所述数据文件已经损毁。

[0066] 在本发明的实施例中,所述ANN可包括具有66至200个隐藏层的神经网络。所述ANN最初可以使用无监督学习方法来训练,因而训练的数据组包括与已知的恶意软件、漏洞及/或网络攻击相关的RREF矩阵组、及/或不受恶意软件影响的RREF矩阵组。所述RREF矩阵组可以根据其个别的性质被标示,例如但不限于依照各自时间范围的读取/写入操作、元数据、存取层级执行绪、硬件唯一识别ID、所耗用的内存、I/O丛发(I/O bursts)、物理位址或软件校验和。由于所述ANN是使用无监督学习方法来训练的,因此经训练的ANN将会在新的威胁被识别时,不断学习这些威胁。在本发明的进一步实施例中,所述ANN可被训练以识别由于特定类型的恶意攻击而分裂的RREF矩阵组的参数。一旦这些参数被识别并且关联到恶意攻击的类型,所述经训练的ANN接着将会能够在未来侦测此种攻击。

[0067] 在本发明的进一步实施例中,所述ANN的神经架构是使用已知病毒类型的100个重要的种类的训练样本数据而被设计的。受影响的特征是根据输入节点的数目、隐藏层的数目、以及在每一个隐藏层中的节点数目而被识别。所述ANN的学习速率一开始参数化为0.01的值,然后随着批次大小为1000逐渐增加。训练周期数从500的值开始。结果显示,当使用800的批次大小和250的训练周期时,显示出最佳结果。

[0068] 经训练的ANN模块124接着继续使用所述经训练的ANN以判断所述更改的RREF矩阵组是否包括恶意活动。若所述经训练的ANN判断所述更改的RREF矩阵组并不包括任何恶意软件,则ANN模块124接着将会使得转换模块122重建所述更改的RREF矩阵组回到其原始的数据文件,并且将其标记为安全的。这些安全的数据文件接着可被回传到主机110作为数据文件112,并且所述映射将会被保持在内存130中。

[0069] 在本发明的实施例中,若所述经训练的ANN判断所述更改的RREF矩阵组包括恶意活动,则将所述更改的RREF矩阵组以及其被映射的更改的数据文件接着移到内存130中的虚拟安全区域。在本发明的实施例中,此虚拟安全区域可以是在内存130中的虚拟沙盒(sandbox)。

[0070] 聚类模块128接着被设置,以使用反向传播模块126从内存130提取更改的RREF矩阵组的变体。在本发明的实施例中,这些变体可包括在一系列的过去的时间范围上的更改的RREF矩阵组、或具有非常类似的值及/或数值的设置的更改的RREF矩阵组。

[0071] 聚类模块128接着将所述更改的RREF矩阵组的这些变体加入内存130中的所述虚拟安全区域。所述更改的RREF矩阵组的所述变体与所述更改的RREF矩阵组一起随后被聚

类。在本发明的实施例中,k-均值聚类函数(k-means clustering function)被用来聚类这些RREF矩阵。

[0072] 聚类模块128随后识别与被发现有恶意活动的所述更改的RREF矩阵组相关的RREF矩阵组。这些先前未被识别的RREF矩阵组随即被标记为潜在可能包括恶意活动的RREF矩阵组,并且随即被系统120保护,使得它们不会不慎地被主机110安装或执行。不与所述更改的RREF矩阵组相关的RREF矩阵组与其对应的被映射的数据文件一起接着被分类为安全的档案。

[0073] 在本发明的实施例中,被发现有恶意活动的RREF矩阵组被聚类模块128提供至经训练的ANN模块124。在本发明的一个实施例中,此信息被使用以进一步训练ANN模块124的经训练的ANN,使得其可以更准确地理解并且在未来识别此种类似的恶意软件。在本发明的另一实施例中,经训练的ANN模块124被用来识别与这些RREF矩阵组相关的恶意活动的类型及/或恶意软件。此信息随后被提供至聚类模块128,使得模块128可以修改其方法,以保护对应于这些RREF矩阵组的数据文件。例如,若经训练的ANN模块124识别之前未发现的RREF矩阵的聚类是与后门计算攻击相关的,则经训练的ANN模块124将会提供此信息至聚类模块128。在接收到此信息时,聚类模块128(其本身或与系统120的其它模块一起)接着可以禁止有关与此聚类的RREF矩阵相关的数据文件的所有进入的数据传输,以避免所述后门计算攻击的发生(亦即通过拒绝从外部的第三方存取这些档案)。

[0074] 在本发明的另一实施例中,ANN模块124将会引入锁密钥机制,以保护被发现是与恶意活动相关的RREF矩阵的最低位元,以确保这些RREF矩阵不会轻易地被恶意的第三方更改,借此禁用其对应的数据文件的漏洞。

[0075] 根据本发明的实施例,图2示出代表可被设置在所述模块的任一个之内以用于实施根据本发明的实施例的处理系统200的部件的框图。本领域技术人员将理解,设置在这些模块之内的每一个处理系统的确切设置可以是不同的,因而处理系统200的确切设置可以改变。因此,图2仅作为示例性说明。

[0076] 在本发明的实施例中,所述模块的每一个可包括控制器201以及用户介面202。用户介面202被设置,以根据需要提供在用户与这些模块的每一个之间的人工的互动,并且为此目的包含所述用户输入指令,以提供更新给这些模块的每一个所需的输入/输出部件。本领域技术人员理解,用户介面202的部件可以随实施例不同而改变,但是通常可包含显示器240、键盘235及轨迹板/触控器236中的一或多个、或者相反地可以不包含这些部件的任一个。

[0077] 控制器201通过汇流排215与用户介面202数据通讯,并且包含安装在电路板之上的内存220、处理器205、操作系统206、用于与用户介面202通讯的输入/输出(I/O)介面230、以及通讯介面。内存220及处理器205处理用于执行此实施例的方法的指令及数据。介面230在此实施例中具有网络卡250的形式。网络卡250例如可被使用以从这些模块经由有线或无线的网络传送数据至其它处理装置、或经由所述有线或无线的网络来接收数据。可被网络卡250使用的无线网络包含但不限于无线网络(Wi-Fi)、蓝牙、近场通讯(NFC)、蜂窝网络、卫星网络、电信网络、广域网络(WAN)等等。

[0078] 内存220及操作系统206通过汇流排210与CPU 205数据通讯。所述内存部件包含易失性及非易失性内存两者、以及每种类型的内存中的超过一种,其包含随机存取内存(RAM)

223、只读内存(ROM)225及储存内存(未显示),而最后一种是包括一个或多个固态硬盘(SSD)。内存220亦包含安全存储器246以用于安全地储存秘密密钥或私密密钥。本领域技术人员理解,上述的内存部件包括非暂态的电脑可读取介质,因而应该被视为包括除了暂态传播信号之外的所有电脑可读取的介质。通常,所述指令是以程序码被储存在所述内存部件中,但亦可以是固线的。内存220可包含核心模块及/或程序化模块,例如是软件应用程序,其可被储存在易失性或非易失性内存中。

[0079] 在此,所述术语"处理器"被用来上位地指任何可以处理此种指令的装置或部件,并且可包含:微处理器、微控制器、可程序化逻辑设备或其它计算设备。换言之,处理器205可以是由任何适当的逻辑电路提供,以用于接收输入、根据储存在内存中的指令来处理它们、以及产生输出(例如产生至所述内存部件、或产生在显示器240上)。在此实施例中,处理器205可以是具有内存可定址空间的单核或多核处理器。在一个示例中,处理器205可以是多核的,例如包括一个8核CPU。在另一示例中,其可以是平行操作以加速计算的CPU核集群。

[0080] 图3描绘根据本发明实施例的用于将从主机获得的一系列的数据文件映射至其对应的RREF矩阵组的方法300。方法300可被实施于系统120中。在系统120通讯地连线于主机110的情况下,系统120能够捕捉在主机110的汇流排系统上发送的所有数据,或储存在主机110的内存中的所有数据。

[0081] 方法300在步骤302从储存在所述主机之内、或通过所述主机发送/接收的所述一系列的数据文件选择数据文件。方法300接着在步骤304将所述数据文件转换为RREF矩阵组。在步骤306,方法300储存所述RREF矩阵组,并且将所述RREF矩阵组映射至所述数据文件。

[0082] 方法300在步骤308继续选择另一数据文件。所述数据文件接着在步骤310被转换为另一RREF矩阵组。在步骤312,方法300储存所述另一RREF矩阵组,并且将所述RREF矩阵组映射至所述另一数据文件。方法300通过自循环步骤314返回执行步骤308,选择另一数据文件、转换所述数据文件以及映射所述数据文件,直到所述一系列的数据文件中的所有数据文件都通过方法300处理为止。

[0083] 图4描绘根据本发明的实施例的、用于保护从主机获得的数据文件的方法400。方法400可以通过系统120的模块实施。方法400在步骤402从所述主机提取更改的数据文件(即第一数据文件)。此更改的数据文件可包括通过方法300储存及处理的数据文件(如图3所描述)的变体。方法400接着在步骤404将所述第一数据文件转换为第一RREF矩阵组。在步骤406,方法400判断所述第一RREF矩阵组是否包含恶意软件或恶意活动。若方法400判断所述第一RREF矩阵组不包括恶意软件,则方法400继续执行步骤408。在步骤408,所述第一数据文件及其对应的RREF矩阵组与其对应的映射一起被储存。

[0084] 相反地,若方法400在步骤406判断所述第一RREF矩阵组包含恶意软件或恶意活动,或与恶意活动相关,则方法400继续执行步骤410,将所述第一RREF矩阵组以及其映射的第一数据文件移到虚拟安全区域。方法400接着在步骤412继续从内存提取所述第一RREF矩阵组的所有变体。所述第一RREF矩阵组的所述变体以及所述第一RREF矩阵组接着通过步骤414执行聚类根据步骤414的聚类结果,方法400在步骤416识别所述第一RREF矩阵组(其先前被视为安全的)的、包含恶意活动的变体,并且亦识别所述第一RREF矩阵组的、并不包含恶意活动的变体。所述第一数据文件接着通过方法400的步骤418,根据不包含恶意活动的

第一RREF矩阵组而被恢复。在步骤420,所述第一RREF矩阵组的包含恶意活动的变体接着由方法400提供至所述经训练的ANN,以用于如前文所述的进一步处理。

[0085] 许多其它的改变、替代、变体及修改可被本领域技术人员所确定,因而所欲的是本发明涵盖所有如同落入所附请求项的范畴之内的此种改变、替代、变体及修改。

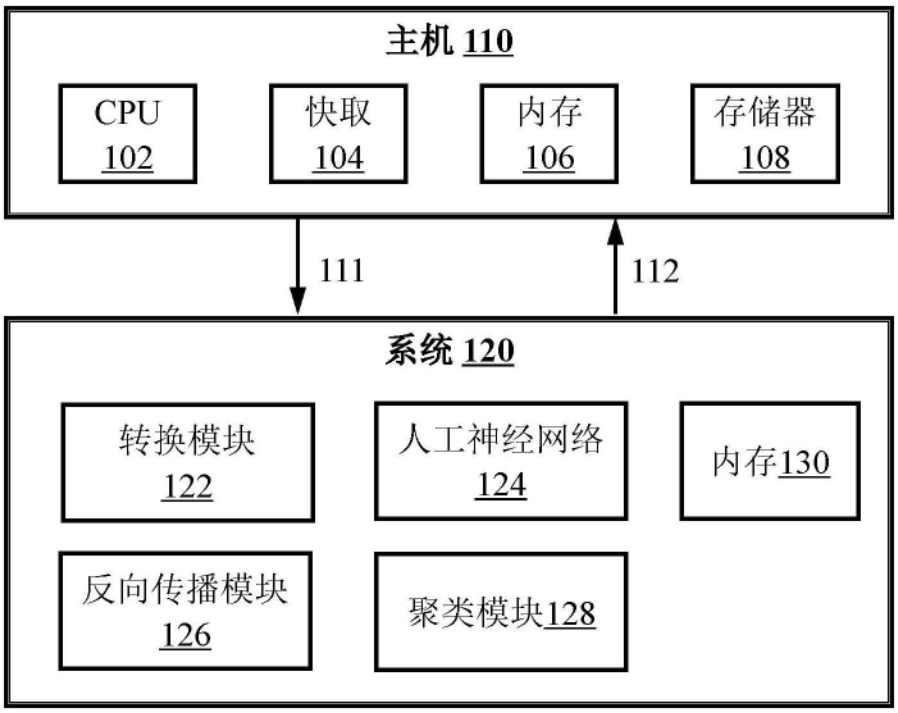

主机 110
CPU 102
快取 104
内存 106
存储器 108
111
112
系统 120
转换模块 122
人工神经网络 124
内存130
反向传播模块 126
聚类模块128

图1

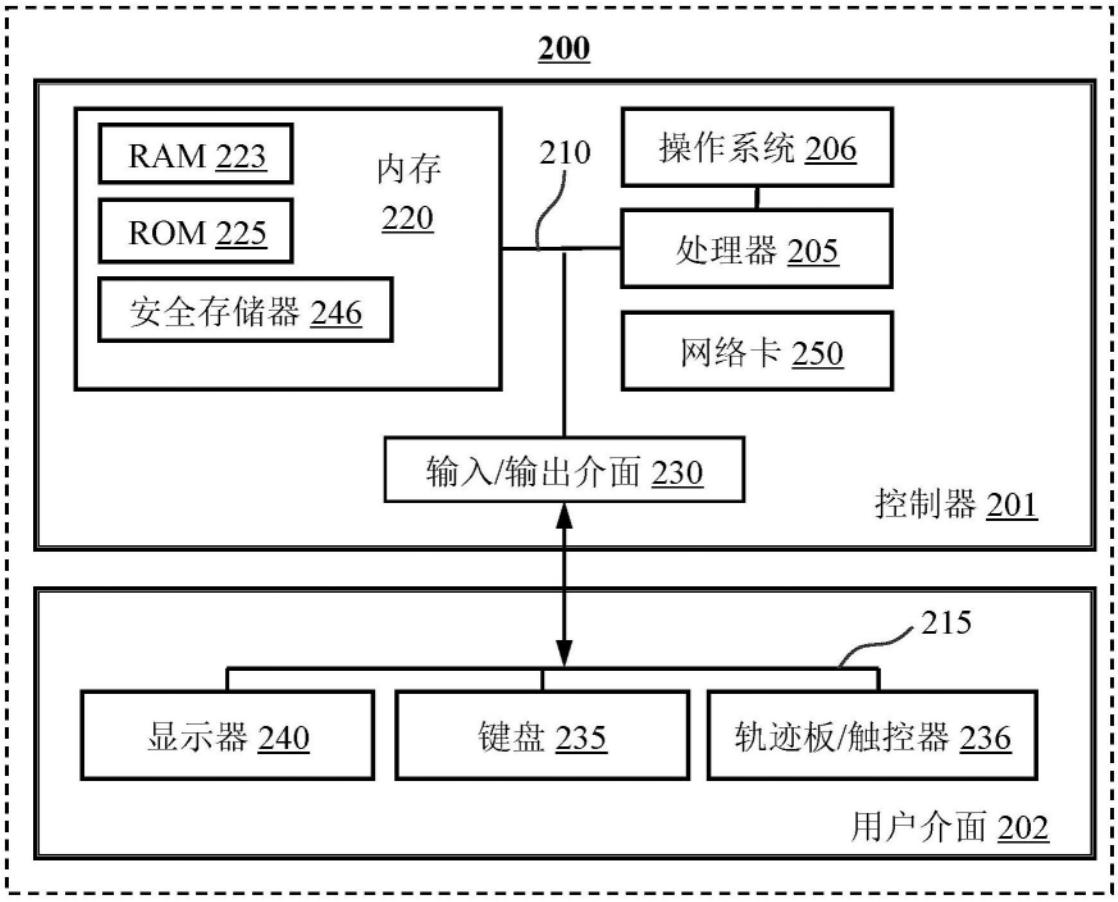

200
RAM 223
ROM 225
安全存储器 246
内存
220
210
操作系统 206
处理器 205
网络卡 250
输入/输出介面 230
控制器 201
215
显示器 240
键盘 235
轨迹板/触控器 236
用户介面 202

图2

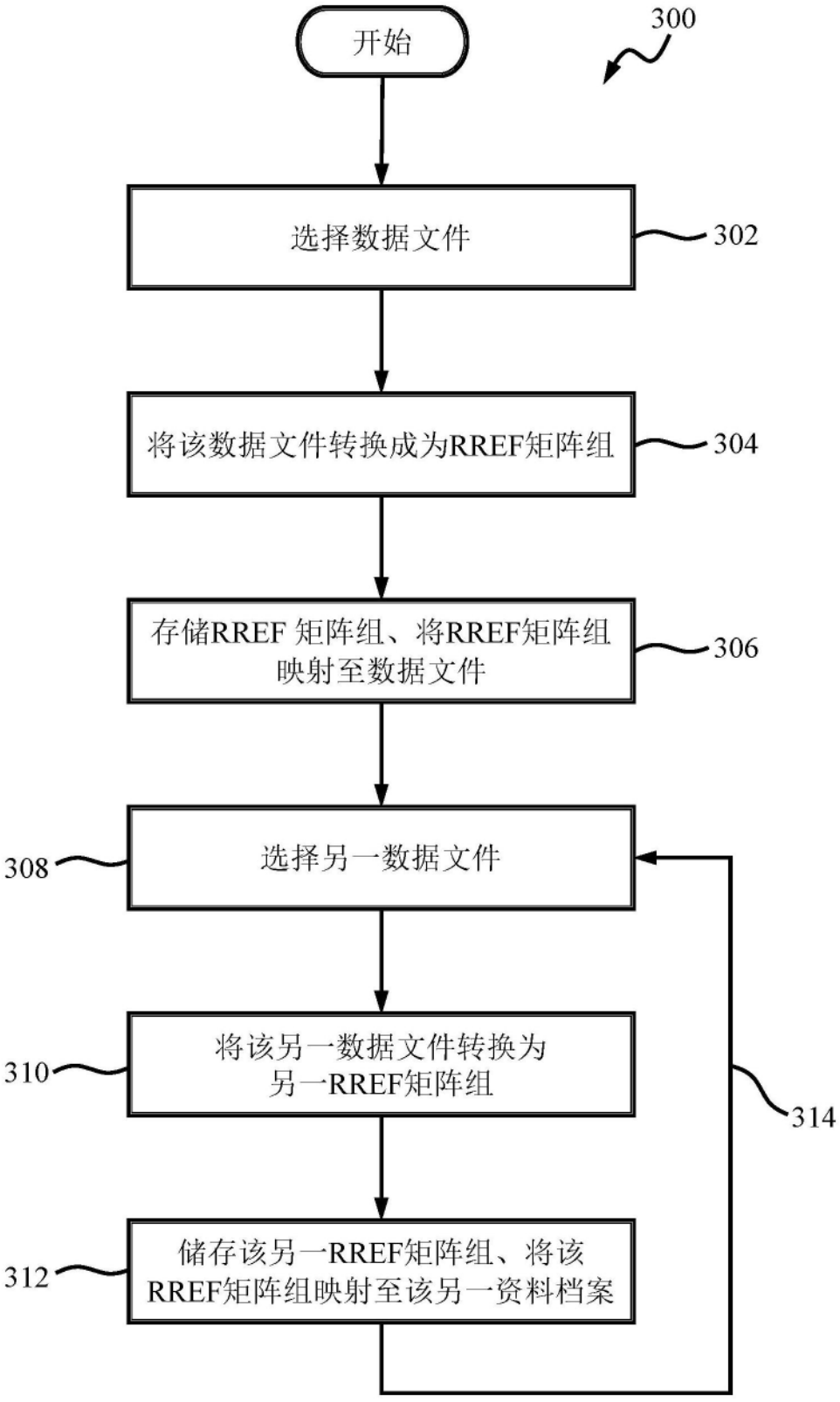

开始
300
选择数据文件
302
将该数据文件转换成为RREF矩阵组
304
存储RREF 矩阵组、将RREF矩阵组映射至数据文件
306
选择另一数据文件
308
将该另一数据文件转换为另一RREF矩阵组
310
储存该另一RREF矩阵组、将该RREF矩阵组映射至该另一资料档案
312
314

图3

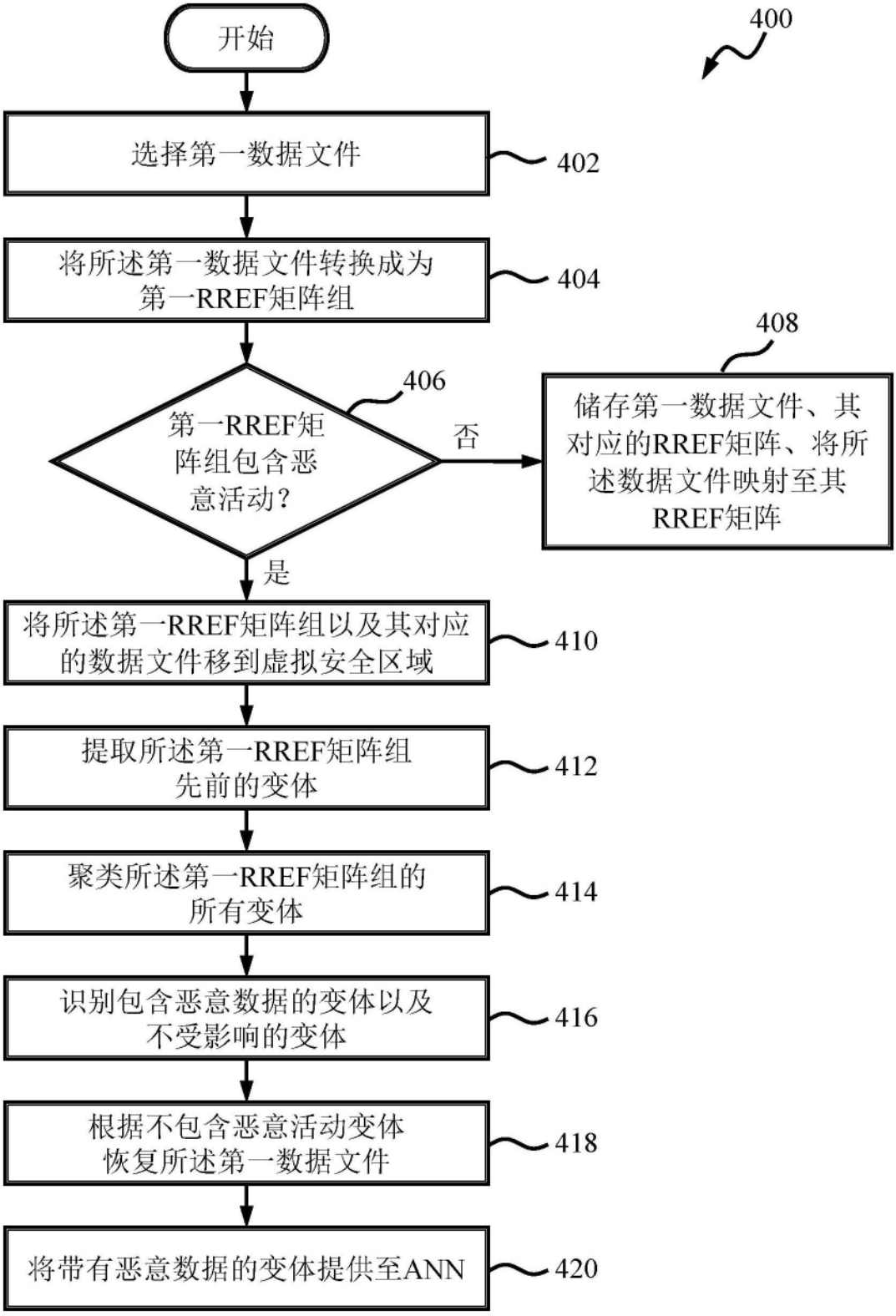

400
开始
选择第一数据文件
402
将所述第一数据文件转换成为第一RREF矩阵组
404
406
第一RREF矩阵组包含恶意活动？
否
408
储存第一数据文件、其对应的RREF矩阵、将所述数据文件映射至其RREF矩阵
是
将所述第一RREF矩阵组以及其对应的数据文件移到虚拟安全区域
410
提取所述第一RREF矩阵组先前的变体
412
聚类所述第一RREF矩阵组的所有变体
414
识别包含恶意数据的变体以及不受影响的变体
416
根据不包含恶意活动变体恢复所述第一数据文件
418
将带有恶意数据的变体提供至ANN
420

图4

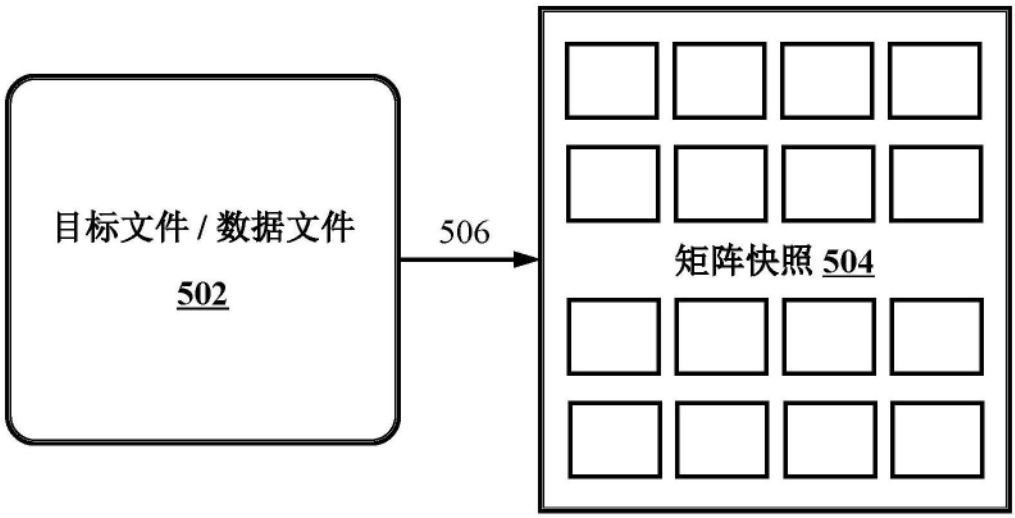

目标文件 / 数据文件
502
506
矩阵快照 504

图5

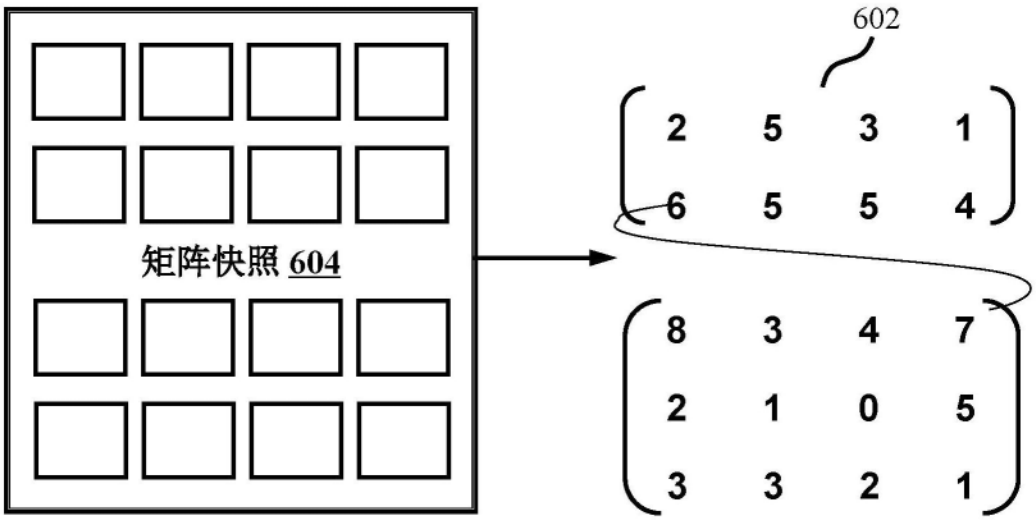

矩阵快照 604
602
2 5 3 1
6 5 5 4
8 3 4 7
2 1 0 5
3 3 2 1

图6