(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113242258 B
(45)授权公告日 2023.11.14

(21)申请号 202110582199.3

(22)申请日 2021.05.27

(65)同一申请的已公布的文献号
申请公布号 CN 113242258 A

(43)申请公布日 2021.08.10

(73)专利权人 安天科技集团股份有限公司
地址 150028 黑龙江省哈尔滨市高新技术产业开发区科技创新城创新创业广场7号楼(世坤路838号)

(72)发明人 黄磊 童志明 肖新光

(74)专利代理机构 北京格允知识产权代理有限公司 11609
专利代理师 张沫

(51)Int.Cl.
H04L 9/40(2022.01)
H04L 67/10(2022.01)
H04L 41/16(2022.01)
G06N 5/01(2023.01)
G06F 18/243(2023.01)

(56)对比文件
CN 101431416 A,2009.05.13
CN 102013992 A,2011.04.13
CN 106657019 A,2017.05.10
CN 109714183 A,2019.05.03
CN 110929886 A,2020.03.27
US 2021089835 A1,2021.03.25
WO 2018095098 A1,2018.05.31
CN 112668913 A,2021.04.16
CN 107835201 A,2018.03.23
CN 111988327 A,2020.11.24
CN 111447215 A,2020.07.24
CN 104468632 A,2015.03.25
CN 105991343 A,2016.10.05
CN 112134854 A,2020.12.25
CN 111181911 A,2020.05.19
CN 108881129 A,2018.11.23
CN 112667651 A,2021.04.16
CN 108021982 A,2018.05.11
CN 109039863 A,2018.12.18
崔阿军;付嘉渝;王玮;闫晓斌;陈力.基于威胁分析技术的网络危险信息源检测方法研究.电子设计工程.2020,(第13期),全文.
侯艳芳;王锦华.基于自更新威胁情报库的大数据安全分析方法.电信科学.2018,(第03期),全文.
(续)

审查员 石琪琦

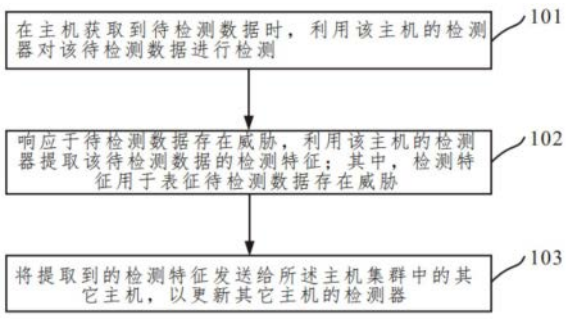
权利要求书2页 说明书9页 附图3页

(54)发明名称

一种主机集群的威胁检测方法和装置

(57)摘要

本发明涉及一种主机集群的威胁检测方法和装置,主机集群包括多个主机,每个主机均包括检测器,主机用于获取待检测数据,该方法包括:在主机获取到待检测数据时,利用该主机的检测器对该待检测数据进行检测;响应于待检测数据存在威胁,利用该主机的检测器提取该待检测数据的检测特征;其中,检测特征用于表征待检测数据存在威胁;将提取到的检测特征发送给主机集群中的其它主机,以更新其它主机的检测器。本方案能够实现主机集群无中心化的威胁检测。



CN 113242258 B

[转续页]

[接上页]

(56)对比文件

刘军等.Ad Hoc网络的一种入侵检测模型.东北大学学报(自然科学版).2006,(第07期),全文.

1.一种主机集群的威胁检测方法,其特征在于,所述主机集群包括多个主机,每个所述主机均包括检测器,所述主机用于获取待检测数据,所述方法包括:

构建检测器资源库;其中,所述检测器资源库包括多个不同类型和不同量级的检测器;

根据每个主机的剩余硬件资源和业务类型,在所述检测器资源库中分别确定出每个主机的检测器的量级和类型;其中,Web服务对应信标检测器,文件存储对应载荷特征检测器,数据库对应数据库检测器;

在主机获取到待检测数据时,利用该主机的检测器对该待检测数据进行检测;

响应于所述待检测数据存在威胁,根据所述待检测数据的类型和该主机的检测器的类型,利用该主机的检测器提取该待检测数据的检测特征;其中,所述检测特征用于表征所述待检测数据存在威胁,当主机确定PE类型的待检测对象存在威胁,针对信标检测器,直接提取全文md5作为其检测特征,针对载荷特征检测器,其提取PE结构中代码节的hash作为其检测特征;

将提取到的检测特征发送给所述主机集群中的其它主机,以更新所述其它主机的检测器;其中,提取到的检测特征优先分发给同类型的检测器,再分发给其它类型的检测器;

所述利用该主机的检测器对该待检测数据进行检测,包括:

利用该主机的检测器中预先构建好的特征库对该待检测数据进行检测;

响应于所述待检测数据不存在威胁,则利用该主机的检测器中预先构建好的威胁检测模型对该待检测数据进行检测;

所述威胁检测模型是通过如下方式进行构建的:

获取历史数据;其中,所述历史数据为在所述待检测数据产生的时间点之前的预设时长的数据;

对所述历史数据进行提取,得到多个特征属性的特征向量;其中,所述特征属性包括:结构特征、API调用特征、传播特征、代码切片特征、签名特征和内容关键字特征;

获取针对每个特征属性的特征向量的标签属性;其中,所述标签属性包括威胁标签和非威胁标签;

将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练,得到威胁检测模型;

所述将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练,得到威胁检测模型,包括:

获取针对每个特征属性的特征向量对应的标签属性赋予的权重;

将得到的每个特征属性的特征向量、该特征向量对应的标签属性和该标签属性所赋予的权重作为训练集对决策树模型进行训练,得到威胁检测模型;其中,不同特征属性的特征向量对应的决策树模型不同。

2.根据权利要求1所述的方法,其特征在于,所述更新所述其它主机的检测器,包括:

更新所述其它主机的检测器的特征库和/或威胁检测模型的权重。

3.一种主机集群的威胁检测装置,其特征在于,所述主机集群包括多个主机,每个所述主机均包括检测器,所述主机用于获取待检测数据,所述装置包括:

部署模块,该部署模块用于执行如下操作:构建检测器资源库;其中,检测器资源库包括多个不同类型和不同量级的检测器;根据每个主机的剩余硬件资源和业务类型,在检测

器资源库中确定出每个主机的检测器的量级和类型;其中,Web服务对应信标检测器,文件存储对应载荷特征检测器,数据库对应数据库检测器;

检测模块,用于在主机获取到待检测数据时,利用该主机的检测器对该待检测数据进行检测;

特征提取模块,用于响应于所述检测模块检测到所述待检测数据存在威胁,根据待检测数据的类型和该主机的检测器的类型,利用该主机的检测器提取该待检测数据的检测特征;其中,所述检测特征用于表征所述待检测数据存在威胁,当主机确定PE类型的待检测对象存在威胁,针对信标检测器,直接提取全文md5作为其检测特征,针对载荷特征检测器,其提取PE结构中代码节的hash作为其检测特征;

分发模块,用于将由所述特征提取模块提取到的检测特征发送给所述主机集群中的其它主机,以更新所述其它主机的检测器;其中,提取到的检测特征优先分发给同类型的检测器,再分发给其它类型的检测器;

所述检测模块具体用于执行如下操作:

利用该主机的检测器中预先构建好的特征库对该待检测数据进行检测;

响应于待检测对象不存在威胁,则利用该主机的检测器中预先构建好的威胁检测模型对该待检测对象进行检测;

进一步包括:模型构建模块,所述模型构建模块用于执行如下操作:

获取历史数据;其中,历史数据为在待检测数据产生的时间点之前的预设时长的数据;

对历史数据进行提取,得到多个特征属性的特征向量;其中,特征属性包括:结构特征、API调用特征、传播特征、代码切片特征、签名特征和内容关键字特征;

获取针对每个特征属性的特征向量的标签属性;其中,标签属性包括威胁标签和非威胁标签;

获取针对每个特征属性的特征向量对应的标签属性赋予的权重;

将得到的每个特征属性的特征向量、该特征向量对应的标签属性和该标签属性所赋予的权重作为训练集对决策树模型进行训练,得到威胁检测模型;其中,不同特征属性的特征向量对应的决策树模型不同。

4.一种主机集群的威胁检测设备,其特征在于,包括:至少一个存储器和至少一个处理器;

所述至少一个存储器,用于存储机器可读程序;

所述至少一个处理器,用于调用所述机器可读程序,执行权利要求1至2中任一项所述的方法。

5.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有计算机指令,所述计算机指令在被处理器执行时,使所述处理器执行权利要求1至2中任一项所述的方法。

一种主机集群的威胁检测方法和装置

技术领域

[0001] 本发明涉及网络安全技术领域，尤其涉及一种主机集群的威胁检测方法和装置。

背景技术

[0002] 目前，针对主机集群的网络威胁检测，一般分为集中式和分布式两种机制，对于集中式检测机制，一旦攻击者成功劫持中心检测器，则会导致整个检测系统失灵；对于分布式检测机制，虽然各检测器独立运行，但每一个检测器均需携带完整的威胁特征库，该威胁特征库由管理中心进行管理和升级，具有中心化特征管理中心，当威胁特征库遭受攻击后，也可能导致整个检测系统失灵，因此针对主机集群的现有检测均无法实现完全无中心化的检测。

[0003] 鉴于此，针对以上不足，需要提供一种主机集群的威胁检测方法和装置来解决上述不足。

发明内容

[0004] 本发明要解决的技术问题在于如何实现主机集群无中心化的威胁检测，针对现有技术中的缺陷，提供了一种主机集群的威胁检测方法和装置。

[0005] 为了解决上述技术问题，第一方面，本发明提供了一种主机集群的威胁检测方法，主机集群包括多个主机，每个主机均包括检测器，主机用于获取待检测数据，该方法包括：

[0006] 在主机获取到待检测数据时，利用该主机的检测器对该待检测数据进行检测；

[0007] 响应于所述待检测数据存在威胁，利用该主机的检测器提取该待检测数据的检测特征；其中，所述检测特征用于表征所述待检测数据存在威胁；

[0008] 将提取到的检测特征发送给所述主机集群中的其它主机，以更新所述其它主机的检测器。

[0009] 可选地，

[0010] 所述利用该主机的检测器对该待检测对象进行检测，包括：

[0011] 利用该主机的检测器中预先构建好的特征库对该待检测对象进行检测；

[0012] 响应于所述待检测对象不存在威胁，则利用该主机的检测器中预先构建好的威胁检测模型对该待检测对象进行检测。

[0013] 可选地，

[0014] 所述威胁检测模型是通过如下方式进行构建的：

[0015] 获取历史数据；其中，所述历史数据为在所述待检测数据产生的时间点之前的预设时长的数据；

[0016] 对所述历史数据进行提取，得到多个特征属性的特征向量；其中，所述特征属性包括如下中的至少一种：结构特征、API调用特征、传播特征、代码切片特征、签名特征和内容关键字特征；

[0017] 获取针对每个特征属性的特征向量的标签属性；其中，所述标签属性包括威胁标

签和非威胁标签；

[0018] 将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练，得到威胁检测模型。

[0019] 可选地，

[0020] 所述将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练，得到威胁检测模型，包括：

[0021] 获取针对每个特征属性的特征向量对应的标签属性赋予的权重；

[0022] 将得到的每个特征属性的特征向量、该特征向量对应的标签属性和该标签属性所赋予的权重作为训练集对决策树模型进行训练，得到威胁检测模型；其中，不同特征属性的特征向量对应的决策树模型不同。

[0023] 可选地，

[0024] 所述更新所述其它主机的检测器，包括：

[0025] 更新所述其它主机的检测器的特征库和/或威胁检测模型的权重。

[0026] 可选地，

[0027] 在所述在主机获取到待检测对象时，利用该主机的检测器对该待检测对象进行检测之前，还包括：

[0028] 构建检测器资源库；其中，所述检测器资源库包括多个不同类型和不同量级的检测器；

[0029] 根据每个主机的剩余硬件资源和/或业务类型，在所述检测器资源库中确定出每个主机的检测器的类型和量级。

[0030] 可选地，

[0031] 所述利用该主机的检测器提取该待检测对象的检测特征，包括：

[0032] 根据所述待检测对象的类型和该主机的检测器的类型，利用该主机的检测器提取该待检测对象的检测特征。

[0033] 第二方面，本发明还提供了一种主机集群的威胁检测装置，所述主机集群包括多个主机，每个所述主机均包括检测器，所述主机用于获取待检测数据，所述装置包括：

[0034] 检测模块，用于在主机获取到待检测数据时，利用该主机的检测器对该待检测数据进行检测；

[0035] 特征提取模块，用于响应于所述检测模块检测到所述待检测数据存在威胁，利用该主机的检测器提取该待检测数据的检测特征；其中，所述检测特征用于表征所述待检测数据存在威胁；

[0036] 分发模块，用于将由所述特征提取模块提取到的检测特征发送给所述主机集群中的其它主机，以更新所述其它主机的检测器。

[0037] 第三方面，本发明还提供了一种主机集群的威胁检测设备，包括：至少一个存储器和至少一个处理器；

[0038] 所述至少一个存储器，用于存储机器可读程序；

[0039] 所述至少一个处理器，用于调用所述机器可读程序，执行上述第一方面或第一方面的任一可能的实现方式所提供的主机集群的威胁检测方法。

[0040] 第四方面，本发明还提供了计算机可读介质，所述计算机可读介质上存储有计算

机指令，所述计算机指令在被处理器执行时，使所述处理器执行上述第一方面或第一方面的任一可能的实现方式所提供的主机集群的威胁检测方法。

[0041] 本发明实施例所提供的一种主机集群的威胁检测方法和装置，在确定该待检测数据存在威胁时，利用该主机的检测器提取该待检测数据的检测特征，并将该检测特征发送给主机集群中的其它主机，以实现其它主机的检测器的更新，从而实现各主机检测器的检测能力的互补和协同检测，进而无需设置特征管理中心以及中心检测器，最终实现完全去中心化的分布式威胁检测。

附图说明

[0042] 图1是本发明实施例所提供的一种主机集群的威胁检测方法；

[0043] 图2是本发明实施例所提供的另一种主机集群的威胁检测方法；

[0044] 图3是本发明实施例所提供的一种主机集群的威胁检测装置所在设备的示意图；

[0045] 图4是本发明实施例所提供的一种主机集群的威胁检测装置的示意图。

具体实施方式

[0046] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明的一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其它实施例，都属于本发明保护的范围。

[0047] 如图1所示，本发明实施例提供的一种主机集群的威胁检测方法，主机集群包括多个主机，每个主机均包括检测器，主机用于获取待检测数据，该方法包括如下步骤：

[0048] 步骤101：在主机获取到待检测数据时，利用该主机的检测器对该待检测数据进行检测；

[0049] 步骤102：响应于待检测数据存在威胁，利用该主机的检测器提取该待检测数据的检测特征；其中，检测特征用于表征待检测数据存在威胁；

[0050] 步骤103：将提取到的检测特征发送给所述主机集群中的其它主机，以更新其它主机的检测器。

[0051] 在本发明实施例中，在确定该待检测数据存在威胁时，利用该主机的检测器提取该待检测数据的检测特征，并将该检测特征发送给主机集群中的其它主机，以实现其它主机的检测器的更新，从而实现各主机检测器的检测能力的互补和协同检测，进而无需设置特征管理中心以及中心检测器，可以避免一旦攻击者成功攻击中心检测器或特征管理库，导致整个威胁检测系统失灵的问题，从而提高了威胁检测的安全性和稳定性，最终实现完全去中心化的分布式威胁检测。

[0052] 在一些实施方式中，待检测对象包括但不限于流量数据和文本数据。其中，流量数据为包括有传播特征、签名特征等的数据；文本数据为包括文件结构、文件内嵌的API函数、代码片段、内置的正文内容关键字、属性等的数据。

[0053] 可选地，在图1所示的一种主机集群的威胁检测方法中，步骤101，包括：

[0054] 利用该主机的检测器中预先构建好的特征库对该待检测对象进行检测；

[0055] 响应于待检测对象不存在威胁，则利用该主机的检测器中预先构建好的威胁检测

模型对该待检测对象进行检测。

[0056] 在本发明实施例中，检测器中均包括有预先构建好的特征库和威胁检测模型。当特征库检测到该待检测对象不存在威胁时，再由威胁检测模型对该待检测对象作出进一步检测，如此，即便是在特征库不更新或被攻击的情况下也实现最大化的协同检测(即特征库和威胁检测模型的协同检测)。同时，基于预先构建的特征库，还可以实现对威胁数据的快速检测；而利用检测模型进行检测，可以避免依赖该特征匹配库，对未知威胁数据进行检测，从而提高了威胁检测的能力和告警率。其中，特征库的检测方式可以是采用正则表达式的检测方式，即将带检测对象与预先构建好的正则表达式去匹配，如果匹配到，则将匹配到的特征作为检测特征。

[0057] 可选地，威胁检测模型是通过如下方式进行构建的：

[0058] 获取历史数据；其中，历史数据为在待检测数据产生的时间点之前的预设时长的数据；

[0059] 对历史数据进行提取，得到多个特征属性的特征向量；其中，特征属性包括如下中的至少一种：结构特征、API调用特征、传播特征、代码切片特征、签名特征和内容关键字特征；

[0060] 获取针对每个特征属性的特征向量的标签属性；其中，标签属性包括威胁标签和非威胁标签；

[0061] 将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练，得到威胁检测模型。

[0062] 在本发明实施例中，针对每个主机中的威胁检测模型，构建方式均为：首先获取预设时长的历史数据，该历史数据包括流量数据和文本数据，对历史数据进行提取，分析文件结构和文件内嵌的API函数、代码片段、内置的正文内容关键字、属性，分析流量数据的传播特征和签名特征，得到结构特征、API调用特征、传播特征、代码切片特征、签名特征和内容关键字特征等多个特征属性的特征向量，由于历史数据为已知数据(即已知该数据为存在威胁或不存在威胁)，故可以获取到针对每个特征属性的特征向量的标签属性(威胁标签和非威胁标签)，将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练，得到威胁检测模型。其中，训练集中包括作为输入的每个特征属性的特征向量以及作为输出的每个特征属性的特征向量的标签属性。如此，基于历史数据能够更准确地获取对应当前主机集群的威胁检测模型，从而提高威胁检测精度。

[0063] 可选地，将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练，得到威胁检测模型，包括：

[0064] 获取针对每个特征属性的特征向量对应的标签属性赋予的权重；

[0065] 将得到的每个特征属性的特征向量、该特征向量对应的标签属性和该标签属性所赋予的权重作为训练集对决策树模型进行训练，得到威胁检测模型；其中，不同特征属性的特征向量对应的决策树模型不同。

[0066] 在本发明实施例中，为了进一步提高威胁检测的精度，赋予每个特征属性以权重，如此通过构建改进的决策树模型，使不同的检测器对应构建不同的决策树模型。具体地，获取该特征属性的特征向量对应的标签属性赋予的权重，以利用得到的对应每个特征属性构建的决策树作为训练集的特征向量、该特征向量对应的标签属性和该标签属性所赋予的权

重作为训练集，对决策树模型进行训练，得到威胁检测模型。

[0067] 在本发明实施例中，需要说明的是，不同的检测器所对应的决策树模型不同，即不同检测器对应获得的特征向量不同和/或特征向量对应的标签属性赋予的权重不同，因此在主机集群中，每个主机的检测器可以独立运行，互不干扰，更不易在遭受攻击出现威胁检测系统瘫痪的现象，从而提高了主机集群中的威胁检测告警率。

[0068] 可选地，在图1所示的一种主机集群的威胁检测方法中，更新其它主机的检测器，包括：

[0069] 更新其它主机的检测器的特征库和/或威胁检测模型的权重。

[0070] 在本发明实施例中，在步骤103将提取到的检测特征发送给主机集群中的其它主机之后，对于每个其它主机，均执行：在训练集中增加该检测特征，其中，该检测特征对应的标签属性为威胁标签，利用更新后的该训练集对决策树模型进行训练，以动态调整该主机的威胁检测模型的权重。

[0071] 在本发明实施例中，在步骤103将提取到的检测特征发送给主机集群中的其它主机之后，对于每个其它主机，均执行：在预先构建好的特征库中增加该检测特征，以对该特征库进行更新。如此，实现了其它各主机检测器的优化，提高了每个主机检测器的检测能力，从而进一步提高了威胁检测的准确性。

[0072] 在本发明实施例中，在步骤102中的主机检测到待检测数据存在威胁时，利用该主机的检测器提取该待检测数据的检测特征之后，进一步包括：利用该检测特征对该主机进行特征库和/或威胁检测模型的权重的更新，以实现自身威胁检测模型的优化，进一步提高检测能力。如此，本发明实施例中的每个检测器均具备自我学习能力，可以进行动态地更新，不断优化各个检测器的检测能力，从而进一步提高主机集群的威胁检测能力。

[0073] 可选地，在图1所示的一种主机集群的威胁检测方法中，在步骤101之前，还包括：

[0074] 构建检测器资源库；其中，检测器资源库包括多个不同类型和不同量级的检测器；

[0075] 根据每个主机的剩余硬件资源和/或业务类型，在检测器资源库中确定出每个主机的检测器的类型和量级。

[0076] 在本发明实施例中，为了保证为每个主机配置相匹配的检测器，需要预先构建检测器资源库，以便对应主机集群中的每个主机的剩余硬件资源和/或业务类型，从检测资源库中确定对应每个主机的确定对应检测器的类型和量级，实现检测器的灵活配置。如此，不同主机的业务类型对应不同类型的检测器，不同的剩余硬件资源对应不同量级的检测器，根据主机剩余硬件资源的情况，提高了主机硬件资源的有效利用。同时还基于主机集群中各主机的业务类型量身确定对应的检测器类型，保证检测的同时实现检测器资源地灵活配置，提高了对检测器资源的有效利用。

[0077] 具体地，针对该主机集群中的每个主机，首先获取该主机对应的业务类型，根据该业务类型，从构建的检测器资源库中确定对应该业务类型的检测器的类型；然后根据该主机的剩余硬件资源，确定该主机对应的检测器的量级，以根据该类型和量级从构建的检测器资源库中确定对应要部署在该主机的检测器，最终将确定的检测器部署在该主机上。

[0078] 在一些实施方式中，获取到待检测数据的主机的剩余硬件资源和/或业务类型是通过如下方式获取的：获取每一个主机的硬件资源信息(例如包括CPU信息、内存信息和磁盘资源信息)和服务信息；对获取到的硬件资源信息和服务信息进行解析提取，得到剩余硬

件资源信息(例如包括剩余CPU信息、剩余内存信息和剩余磁盘资源信息)和业务类型(例如包括Web服务类型、数据库服务类型和文件存储服务类型);将得到的剩余硬件资源信息和业务类型进行存储,得到信息库;根据获取到待检测数据的主机的标识信息和得到的信息库,确定该主机的剩余硬件资源和/或业务类型。

[0079] 具体地,根据该主机的剩余硬件资源,确定该主机对应的检测器的量级,包括:判断该主机的剩余硬件资源是否大于预设硬件资源阈值,如果是,则确定对应该主机的检测器的量级为重量级;如果否,则确定对应该主机的检测器的量级为轻量级。

[0080] 例如,主机集群中包括6台主机服务器(编号为1、2、3、4、5、6),根据业务类型可以划分为2台Web服务器(编号为1、2)、2台数据库服务器(编号为3、4)和2台文件存储服务器(编号为5、6)。其中,根据检测器资源库,确定业务类型与检测器类型的对应关系,Web服务对应信标检测器,文件存储对应载荷特征检测器,数据库对应数据库检测器;且剩余硬件资源大于预设硬件资源阈值的为编号为1、3、5的服务器,小于预设硬件资源阈值的为编号为2、4、6的服务器。综上所述,对应编号为1的服务器应部署信标重量级检测器,对应编号为2的服务器应部署信标轻量级检测器,对应编号为3的服务器应部署数据库重量级检测器,对应编号为4的服务器应部署数据库轻量级检测器,对应编号为5的服务器应部署载荷特征重量级检测器,对应编号为6的服务器应部署载荷特征轻量级检测器。

[0081] 可选地,在图1所示的一种主机集群的威胁检测方法中,步骤102,包括:

[0082] 根据待检测对象的类型和该主机的检测器的类型,利用该主机的检测器提取该待检测对象的检测特征。

[0083] 在本发明实施例中,在确定待检测对象存在威胁时,根据待检测对象的类型和该主机的检测器的类型,利用该主机的检测器提取该待检测对象的检测特征。例如,接前例所述,当主机确定PE类型的待检测对象存在威胁,针对信标检测器,直接提取全文md5作为其检测特征,针对载荷特征检测器,其提取PE结构中代码节的hash作为其检测特征。

[0084] 在本发明实施例中,步骤103中检测特征的自动分发是基于专有的加密协议实现的,其中,优先分发给同类型的检测器,再分发给其它类型的检测器,由于威胁攻击具有对相同类型检测器进行攻击的倾向,因此能够优先保证同类型检测器能快速实现针对同类型待检测数据的威胁检测。此外,自动分发所采用的加密协议进一步保证了分发的安全性,能够防止攻击者攻克后发布虚假消息,避免其它主机获取到伪装的分发特征信息,进而确保威胁检测的准确性。

[0085] 为了更加清楚地说明本发明的技术方案及优点,如图2所示,下面对本发明实施例提供的一种主机集群的威胁检测方法进行详细的说明,具体包括:

[0086] 步骤201:构建检测器资源库。

[0087] 步骤202:根据每个主机的剩余硬件资源和/或业务类型,在检测器资源库中确定出每个主机的检测器的类型和量级。

[0088] 步骤203:在主机获取到待检测数据时,利用该主机的检测器中预先构建好的特征库对该待检测对象进行检测。

[0089] 步骤204:响应于待检测对象不存在威胁,则利用该主机的检测器中预先构建好的威胁检测模型对该待检测对象进行检测。

[0090] 步骤205:响应于待检测数据存在威胁,根据待检测对象的类型和该主机的检测器

的类型，利用该主机的检测器提取该待检测对象的检测特征。

[0091] 步骤206：将提取到的检测特征发送给主机集群中的其它主机，以更新其它主机的检测器的特征库和/或威胁检测模型的权重。

[0092] 如图3、图4所示，本发明实施例提供了一种主机集群的威胁检测装置。装置实施例可以通过软件实现，也可以通过硬件或者软硬件结合的方式实现。从硬件层面而言，如图3所示，为本发明实施例提供的一种主机集群的威胁检测装置所在设备的一种硬件结构图，除了图3所示的处理器、内存、网络接口、以及非易失性存储器之外，实施例中装置所在的设备通常还可以包括其它硬件，如负责处理报文的转发芯片等等。以软件实现为例，如图4所示，作为一个逻辑意义上的装置，是通过其所在设备的CPU将非易失性存储器中对应的计算机程序指令读取到内存中运行形成的。本实施例提供的一种主机集群的威胁检测装置，主机集群包括多个主机，每个主机均包括检测器，主机用于获取待检测数据，包括：

[0093] 检测模块401，用于在主机获取到待检测数据时，利用该主机的检测器对该待检测数据进行检测；

[0094] 特征提取模块402，用于响应于检测模块401检测到待检测数据存在威胁，利用该主机的检测器提取该待检测数据的检测特征；其中，检测特征用于表征待检测数据存在威胁；

[0095] 分发模块403，用于将由特征提取模块403提取到的检测特征发送给主机集群中的其它主机，以更新其它主机的检测器。

[0096] 可选地，在图4所示一种主机集群的威胁检测装置的基础上，检测模块401，还用于执行如下操作：

[0097] 利用该主机的检测器中预先构建好的特征库对该待检测对象进行检测；

[0098] 响应于待检测对象不存在威胁，则利用该主机的检测器中预先构建好的威胁检测模型对该待检测对象进行检测。

[0099] 可选地，在图4所示一种主机集群的威胁检测装置的基础上，该装置进一步包括：模型构建模块，该模型构建模块用于执行如下操作：

[0100] 获取历史数据；其中，历史数据为在待检测数据产生的时间点之前的预设时长的数据；

[0101] 对历史数据进行提取，得到多个特征属性的特征向量；其中，特征属性包括如下中的至少一种：结构特征、API调用特征、传播特征、代码切片特征、签名特征和内容关键字特征；

[0102] 获取针对每个特征属性的特征向量的标签属性；其中，标签属性包括威胁标签和非威胁标签；

[0103] 将得到的每个特征属性的特征向量及其对应的标签属性作为训练集对决策树模型进行训练，得到威胁检测模型。

[0104] 可选地，在图4所示一种主机集群的威胁检测装置的基础上，模型构建模块还用于执行如下操作：

[0105] 获取针对每个特征属性的特征向量对应的标签属性赋予的权重；

[0106] 将得到的每个特征属性的特征向量、该特征向量对应的标签属性和该标签属性所赋予的权重作为训练集对决策树模型进行训练，得到威胁检测模型；其中，不同特征属性的

特征向量对应的决策树模型不同。

[0107] 可选地，在图4所示一种主机集群的威胁检测装置的基础上，该装置进一步包括：更新模块，该更新模块用于执行如下操作：

[0108] 更新其它主机的检测器的特征库和/或威胁检测模型的权重。

[0109] 可选地，在图4所示一种主机集群的威胁检测装置的基础上，该装置进一步包括：部署模块，该部署模块用于执行如下操作：

[0110] 构建检测器资源库；其中，检测器资源库包括多个不同类型和不同量级的检测器；

[0111] 根据每个主机的剩余硬件资源和/或业务类型，在检测器资源库中确定出每个主机的检测器的类型和量级。

[0112] 可选地，在图4所示一种主机集群的威胁检测装置的基础上，特征提取模块402还用于执行如下操作：

[0113] 根据待检测对象的类型和该主机的检测器的类型，利用该主机的检测器提取该待检测对象的检测特征。

[0114] 可以理解的是，本发明实施例示意的结构并不构成对一种主机集群的威胁检测装置的具体限定。在本发明的另一些实施例中，一种主机集群的威胁检测装置可以包括比图示更多或者更少的部件，或者组合某些部件，或者拆分某些部件，或者不同的部件布置。图示的部件可以以硬件、软件或者软件和硬件的组合来实现。

[0115] 上述装置内的各模块之间的信息交互、执行过程等内容，由于与本发明方法实施例基于同一构思，具体内容可参见本发明方法实施例中的叙述，此处不再赘述。

[0116] 本发明实施例还提供了一种主机集群的威胁检测设备，包括：至少一个存储区和至少一个处理器；

[0117] 所述至少一个存储器，用于存储机器可读程序；

[0118] 所述至少一个处理器，用于调用所述机器可读程序，执行本发明任一实施例中的一种主机集群的威胁检测方法。

[0119] 本发明实施例还提供了一种计算机可读介质，所述计算机可读介质上存储有计算机指令，所述计算机指令在被处理器执行时，使所述处理器执行本发明任一实施例中的一种主机集群的威胁检测方法。

[0120] 具体地，可以提供配有存储介质的系统或者装置，在该存储介质上存储着实现上述实施例中任一实施例的功能的软件程序代码，且使该系统或者装置的计算机(或CPU或MPU)读出并执行存储在存储介质中的程序代码。

[0121] 在这种情况下，从存储介质读取的程序代码本身可实现上述实施例中任何一项实施例的功能，因此程序代码和存储程序代码的存储介质构成了本发明的一部分。

[0122] 用于提供程序代码的存储介质实施例包括软盘、硬盘、磁光盘、光盘(如CD-ROM、CD-R、CD-RW、DVD-ROM、DVD-RAM、DVD-RW、DVD+RW)、磁带、非易失性存储卡和ROM。可选择地，可以由通信网络从服务器计算机上下载程序代码。

[0123] 此外，应该清楚的是，不仅可以通过执行计算机所读出的程序代码，而且可以通过基于程序代码的指令使计算机上操作的操作系统等来完成部分或者全部的实际操作，从而实现上述实施例中任意一项实施例的功能。

[0124] 此外，可以理解的是，将由存储介质读出的程序代码写到插入计算机内的扩展板

中所设置的存储器中或者写到与计算机相连接的扩展模块中设置的存储器中，随后基于程序代码的指令使安装在扩展板或者扩展模块上的CPU等来执行部分和全部实际操作，从而实现上述实施例中任一实施例的功能。

[0125] 需要说明的是，在本文中，诸如第一和第二之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其它变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其它要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个·····”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同因素。

[0126] 本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述的程序可以存储在计算机可读取的存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质中。

[0127] 最后应说明的是：以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

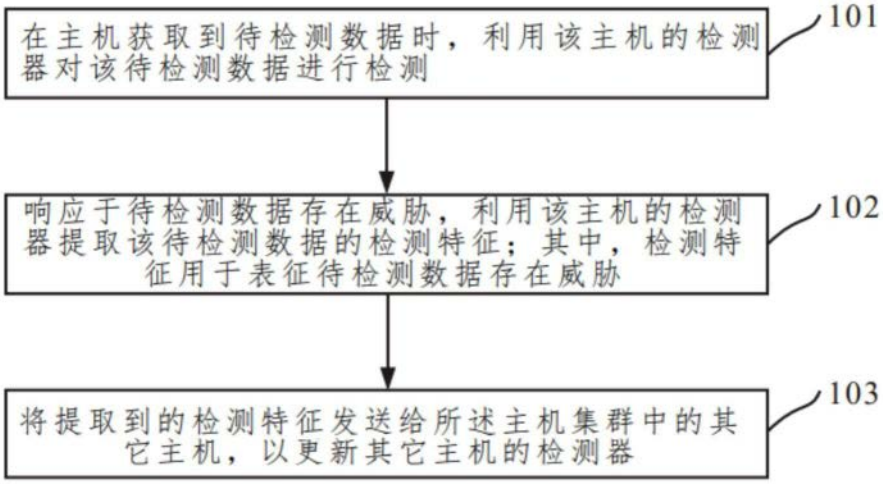

在主机获取到待检测数据时，利用该主机的检测器对该待检测数据进行检测
101
响应于待检测数据存在威胁，利用该主机的检测器提取该待检测数据的检测特征；其中，检测特征用于表征待检测数据存在威胁
102
将提取到的检测特征发送给所述主机集群中的其它主机，以更新其它主机的检测器
103

图1

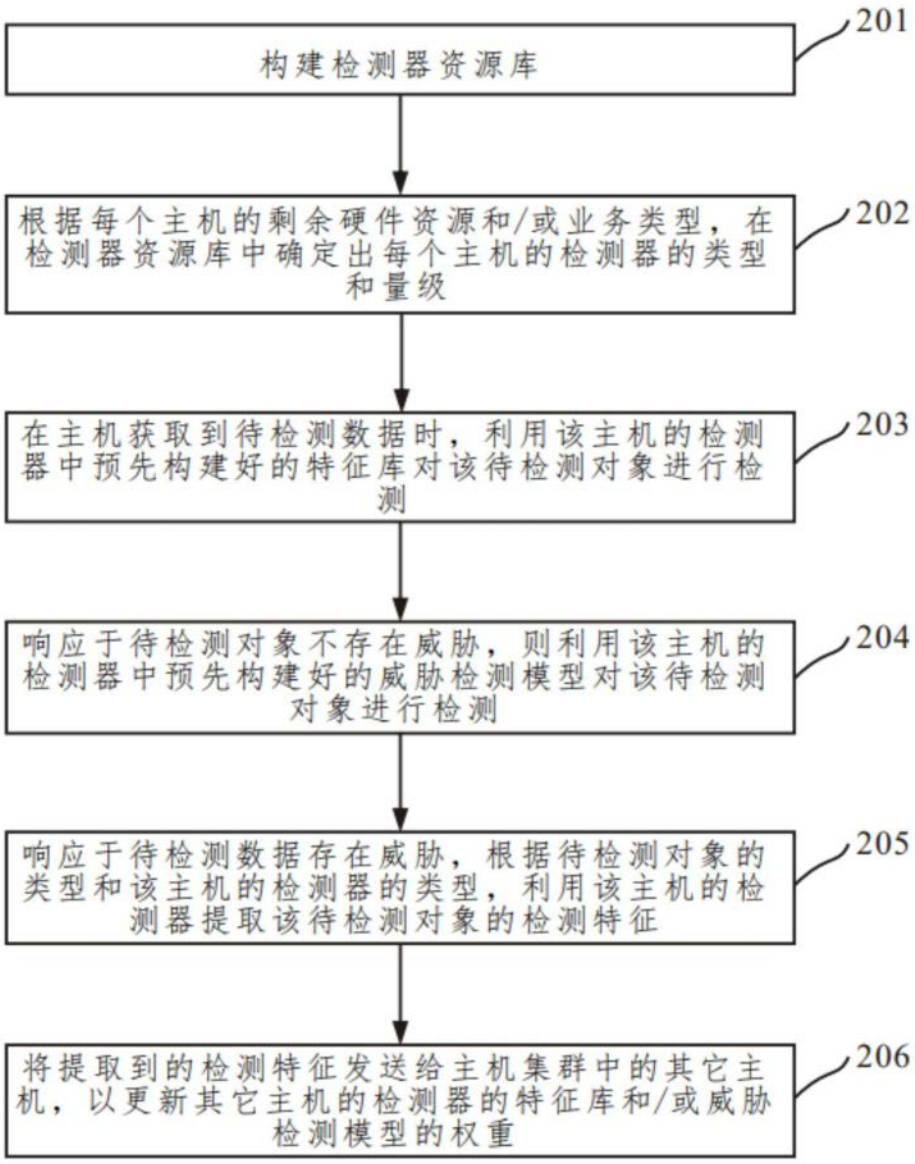

201
构建检测器资源库
202
根据每个主机的剩余硬件资源和/或业务类型，在检测器资源库中确定出每个主机的检测器的类型和量级
203
在主机获取到待检测数据时，利用该主机的检测器中预先构建好的特征库对该待检测对象进行检测
204
响应于待检测对象不存在威胁，则利用该主机的检测器中预先构建好的威胁检测模型对该待检测对象进行检测
205
响应于待检测数据存在威胁，根据待检测对象的类型和该主机的检测器的类型，利用该主机的检测器提取该待检测对象的检测特征
206
将提取到的检测特征发送给主机集群中的其它主机，以更新其它主机的检测器的特征库和/或威胁检测模型的权重

图2

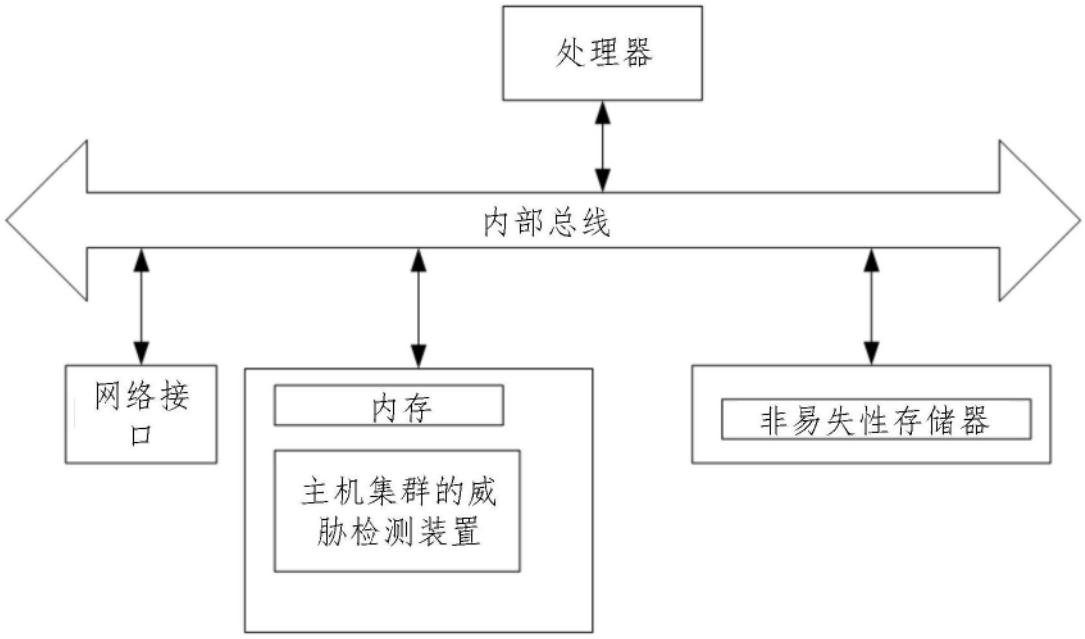

处理器
内部总线
网络接口
内存
主机集群的威胁检测装置
非易失性存储器

图3

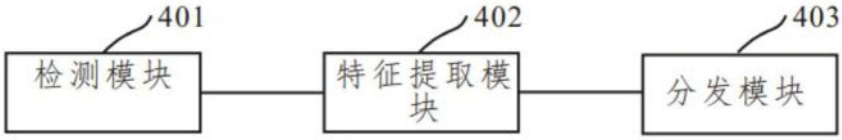

401
检测模块
402
特征提取模块
403
分发模块

图4