(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 111565205 B
(45)授权公告日 2020.10.23

(21)申请号 202010684168.4

(22)申请日 2020.07.16

(65)同一申请的已公布的文献号
申请公布号 CN 111565205 A

(43)申请公布日 2020.08.21

(73)专利权人 腾讯科技(深圳)有限公司
地址 518000 广东省深圳市南山区高新区科技中一路腾讯大厦35层

(72)发明人 邓真 林智鑫 向琦

(74)专利代理机构 广州华进联合专利商标代理有限公司 44224
代理人 李文渊 杨欢

(51)Int.Cl.
H04L 29/06(2006.01)
G06K 9/62(2006.01)

(56)对比文件
CN 111030986 A,2020.04.17
CN 111030986 A,2020.04.17
CN 108446559 A,2018.08.24
CN 106375331 A,2017.02.01
CN 111193749 A,2020.05.22
CN 110837640 A,2020.02.25
US 2010050260 A1,2010.02.25

审查员 杨志忠

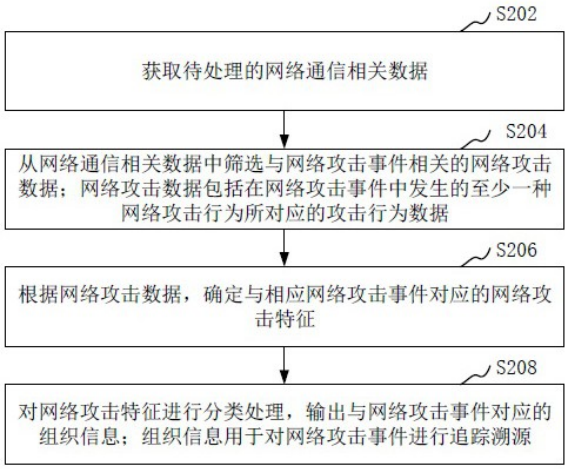
权利要求书4页 说明书20页 附图8页

(54)发明名称

网络攻击识别方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种网络攻击识别方法、装置、计算机设备和存储介质。所述方法包括:获取待处理的网络通信相关数据;从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据;所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据;根据所述网络攻击数据,确定与相应网络攻击事件对应的网络攻击特征;对所述网络攻击特征进行分类处理,输出与所述网络攻击事件对应的组织信息;所述组织信息用于对所述网络攻击事件进行追踪溯源。采用本方法能够提高网络攻击所属组织的识别准确性,保障网络安全。



CN 111565205 B

1.一种网络攻击识别方法,其特征在于,所述方法包括:

获取待处理的网络通信相关数据;

从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据;所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据;

获取威胁建模模型,并根据所述威胁建模模型,将所述网络攻击数据中的攻击行为数据映射成对应的攻击行为特征;

将所述网络攻击数据中的基础设施数据,转换成对应的基础设施特征;

根据所述攻击行为特征和所述基础设施特征,确定与所述网络攻击事件对应的网络攻击特征;

通过检测模型对所述网络攻击特征进行分类处理,输出与所述网络攻击事件对应的组织信息;所述组织信息用于对所述网络攻击事件进行追踪溯源;所述检测模型是通过训练数据集训练得到的具有分类功能的模型,所述训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据。

2.根据权利要求1所述的方法,其特征在于,所述获取待处理的网络通信相关数据,包括:

确定部署于预设区域内的网络设备;所述网络设备包括交换机和主机设备;

获取通过所述交换机进行转发的网络流量数据;

获取所述主机设备在运行时产生的进程数据、线程数据和日志数据;

将所述网络流量数据、所述进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据。

3.根据权利要求1所述的方法,其特征在于,所述从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据,包括:

确定所述网络通信相关数据中与不同网络行为分别对应的网络行为数据;

当所述网络行为数据为攻击行为数据时,确定产生所述攻击行为数据的发起方;

将所述网络通信相关数据中所有与所述发起方相关的网络行为数据、以及所述发起方对应的基础设施数据,作为与网络攻击事件相关的网络攻击数据。

4.根据权利要求3所述的方法,其特征在于,所述当所述网络行为数据为攻击行为数据时,确定产生所述攻击行为数据的发起方之前,所述方法还包括:

对所述网络行为数据进行分析,确定各种网络行为分别发生的频次、以及各种网络行为所对应的行为关键信息;

当所述频次大于等于阈值,或者所述行为关键信息中包括有恶意关键词时,确定相应的网络行为为网络攻击行为;

将与所述网络攻击行为相关的网络行为数据作为攻击行为数据。

5.根据权利要求1所述的方法,其特征在于,所述检测模型包括决策树模型,所述通过检测模型对所述网络攻击特征进行分类处理,输出与所述网络攻击事件对应的组织信息,包括:

获取预先构建的决策树模型;

依据所述网络攻击特征中不同特征维度各自对应的特征值,从所述决策树模型的根节点开始,不断地自上向下从所述决策树模型中查找与所述网络攻击特征相匹配的目标内部

节点，直至达到目标叶子节点时为止；

将所述目标叶子节点中存储的组织信息，作为所述网络攻击事件的发起方所对应的组织信息并输出。

6.根据权利要求5所述的方法，其特征在于，所述依据所述网络攻击特征中不同特征维度各自对应的特征值，从所述决策树模型的根节点开始，不断地自上向下从所述决策树模型中查找与所述网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止，包括：

从所述决策树模型的根节点开始，根据所述网络攻击特征中与所述根节点所对应特征维度的特征值，确定下一层的目标内部节点；

根据所述网络攻击特征中与所述下一层的目标内部节点所对应特征维度的特征值，确定再下一层的目标内部节点，并不断往下查找与所述网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止。

7.根据权利要求1至6中任一项所述的方法，其特征在于，所述检测模型包括决策树模型，所述决策树模型的构建步骤包括：

获取训练数据集；

根据所述样本网络攻击数据，确定与所述样本组织信息对应的样本网络攻击特征；

通过所述样本网络攻击特征和对应的样本组织信息构建决策树模型。

8.根据权利要求7所述的方法，其特征在于，所述样本网络攻击数据包括样本攻击行为数据和样本基础设施数据；所述根据所述样本网络攻击数据，确定与所述样本组织信息对应的样本网络攻击特征，包括：

根据所述威胁建模模型，将所述样本攻击行为数据映射成对应的样本攻击行为特征；

将所述样本基础设施数据转换成对应的样本基础设施特征；

根据所述样本攻击行为特征和所述样本基础设施特征，确定与所述样本组织信息对应的样本网络攻击特征。

9.根据权利要求7所述的方法，其特征在于，所述通过所述样本网络攻击特征和对应的样本组织信息构建决策树模型，包括：

确定与所述样本网络攻击特征对应的多于一个的特征维度；

根据所述训练数据集，从所述样本网络攻击特征对应的多于一个的特征维度中选择其中一个特征维度作为分类特征以创建根节点，并根据选择的分类特征将训练数据集分裂成多个训练子集；

在分裂产生的训练子集中不断的选择分类特征创建内部节点，并根据选择的分类特征进行数据分裂产生新的训练子集，直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上；

根据各样本组织信息创建对应的叶子节点；

根据创建的根节点、所述根节点之下的内部节点、以及所述叶子节点，确定决策树模型。

10.根据权利要求9所述的方法，其特征在于，所述根据所述训练数据集，从所述样本网络攻击特征对应的多于一个的特征维度中选择其中一个特征维度作为分类特征以创建根节点，包括：

根据所述训练数据集计算每个特征维度分别对应的信息增益率；

将所述信息增益率中的最大信息增益率所对应的特征维度，作为与所述训练数据集对应的分类特征；

根据与所述训练数据集对应的分类特征创建根节点。

11.根据权利要求1所述的方法，其特征在于，所述方法还包括：

获取预设时间段内，通过所述检测模型对所述网络攻击特征处理所输出的组织信息；

根据预设时间段内处理的网络攻击特征和相应输出的组织信息，更新所述训练数据集，并基于更新后的训练数据集对所述检测模型进行更新。

12.一种网络攻击识别装置，其特征在于，所述装置包括：

获取模块，用于获取待处理的网络通信相关数据；

筛选模块，用于从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据；

确定模块，用于获取威胁建模模型，并根据所述威胁建模模型，将所述网络攻击数据中的攻击行为数据映射成对应的攻击行为特征，将所述网络攻击数据中的基础设施数据，转换成对应的基础设施特征，并根据所述攻击行为特征和所述基础设施特征，确定与所述网络攻击事件对应的网络攻击特征；

分类模块，用于通过检测模型对所述网络攻击特征进行分类处理，输出与所述网络攻击事件对应的组织信息；所述组织信息用于对所述网络攻击事件进行追踪溯源；所述检测模型是通过训练数据集训练得到的具有分类功能的模型，所述训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据。

13.根据权利要求12所述的装置，其特征在于，所述获取模块，具体用于确定部署于预设区域内的网络设备，其中，所述网络设备包括交换机和主机设备；获取通过所述交换机进行转发的网络流量数据；获取所述主机设备在运行时产生的进程数据、线程数据和日志数据；将所述网络流量数据、所述进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据。

14.根据权利要求12所述的装置，其特征在于，所述筛选模块，具体用于确定所述网络通信相关数据中与不同网络行为分别对应的网络行为数据；当所述网络行为数据为攻击行为数据时，确定产生所述攻击行为数据的发起方；将所述网络通信相关数据中所有与所述发起方相关的网络行为数据、以及所述发起方对应的基础设施数据，作为与网络攻击事件相关的网络攻击数据。

15.根据权利要求14所述的装置，其特征在于，所述筛选模块，还用于对所述网络行为数据进行分析，确定各种网络行为分别发生的频次、以及各种网络行为所对应的行为关键信息；当所述频次大于等于阈值，或者所述行为关键信息中包括有恶意关键词时，确定相应的网络行为为网络攻击行为；将与所述网络攻击行为相关的网络行为数据作为攻击行为数据。

16.根据权利要求12所述的装置，其特征在于，所述检测模型包括决策树模型，所述分类模块，具体用于获取预先构建的决策树模型；依据所述网络攻击特征中不同特征维度各自对应的特征值，从所述决策树模型的根节点开始，不断地自上向下从所述决策树模型中

查找与所述网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止；将所述目标叶子节点中存储的组织信息，作为所述网络攻击事件的发起方所对应的组织信息并输出。

17.根据权利要求16所述的装置，其特征在于，所述分类模块，具体用于从所述决策树模型的根节点开始，根据所述网络攻击特征中与所述根节点所对应特征维度的特征值，确定下一层的目标内部节点；根据所述网络攻击特征中与所述下一层的目标内部节点所对应特征维度的特征值，确定再下一层的目标内部节点，并不断往下查找与所述网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止。

18.根据权利要求12至17中任一项所述的装置，其特征在于，所述检测模型包括决策树模型，所述装置还包括模型构建模块，用于获取训练数据集；根据所述样本网络攻击数据，确定与所述样本组织信息对应的样本网络攻击特征；通过所述样本网络攻击特征和对应的样本组织信息构建决策树模型。

19.根据权利要求18所述的装置，其特征在于，所述模型构建模块具体用于根据所述威胁建模模型，将所述样本攻击行为数据映射成对应的样本攻击行为特征；将所述样本基础设施数据转换成对应的样本基础设施特征；根据所述样本攻击行为特征和所述样本基础设施特征，确定与所述样本组织信息对应的样本网络攻击特征。

20.根据权利要求18所述的装置，其特征在于，所述模型构建模块具体用于确定与所述样本网络攻击特征对应的多于一个的特征维度；根据所述训练数据集，从所述样本网络攻击特征对应的多于一个的特征维度中选择其中一个特征维度作为分类特征以创建根节点，并根据选择的分类特征将训练数据集分裂成多个训练子集；在分裂产生的训练子集中不断的选择分类特征创建内部节点，并根据选择的分类特征进行数据分裂产生新的训练子集，直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上；根据各样本组织信息创建对应的叶子节点；根据创建的根节点、所述根节点之下的内部节点、以及所述叶子节点，确定决策树模型。

21.根据权利要求20所述的装置，其特征在于，所述模型构建模块具体用于根据所述训练数据集计算每个特征维度分别对应的信息增益率；将所述信息增益率中的最大信息增益率所对应的特征维度，作为与所述训练数据集对应的分类特征；根据与所述训练数据集对应的分类特征创建根节点。

22.根据权利要求12所述的装置，其特征在于，所述装置还包括模型更新模块，用于获取预设时间段内，通过所述检测模型对所述网络攻击特征处理所输出的组织信息；根据预设时间段内处理的网络攻击特征和相应输出的组织信息，更新所述训练数据集，并基于更新后的训练数据集对所述检测模型进行更新。

23.一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现权利要求1至11中任一项所述的方法的步骤。

24.一种计算机可读存储介质，存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求1至11中任一项所述的方法的步骤。

网络攻击识别方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及网络安全技术领域，特别是涉及一种网络攻击识别方法、装置、计算机设备和存储介质。

背景技术

[0002] 随着网络技术的发展，出现了网络安全技术，网络安全技术主要是为了维护计算机通信网络的安全，主要包括网络的硬件和软件的正常运行、以及数据信息交换的安全。在实际应用中，由于网络攻击行为的频发常常会对系统的网络安全造成隐患。为了保障系统安全，对网络攻击进行识别，并确定网络攻击发起方所属家族就变得尤为重要了。

[0003] 传统方案中，识别网络攻击行为所属家族的判定方式，通常会对恶意攻击行为所对应的代码特征、恶意文件或恶意程序所对应的哈希特征、以及网络流量特征等进行精准的特征对比，以对网络攻击事件进行家族分类判断。然而，传统的网络攻击所属家族判定方式，由于判定特征以及维度较为单一和固有，无法准确对网络攻击事件进行追踪溯源。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种能够提高对网络攻击事件进行追踪溯源的网络攻击识别方法、装置、计算机设备和存储介质。

[0005] 一种网络攻击识别方法，所述方法包括：

[0006] 获取待处理的网络通信相关数据；

[0007] 从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据；

[0008] 根据所述网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征；

[0009] 对所述网络攻击特征进行分类处理，输出与所述网络攻击事件对应的组织信息；所述组织信息用于对所述网络攻击事件进行追踪溯源。

[0010] 一种网络攻击识别装置，所述装置包括：

[0011] 获取模块，用于获取待处理的网络通信相关数据；

[0012] 筛选模块，用于从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据；

[0013] 确定模块，用于根据所述网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征；

[0014] 分类模块，用于对所述网络攻击特征进行分类处理，输出与所述网络攻击事件对应的组织信息；所述组织信息用于对所述网络攻击事件进行追踪溯源。

[0015] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现以下步骤：

[0016] 获取待处理的网络通信相关数据；

[0017] 从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据；

[0018] 根据所述网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征；

[0019] 对所述网络攻击特征进行分类处理，输出与所述网络攻击事件对应的组织信息；所述组织信息用于对所述网络攻击事件进行追踪溯源。

[0020] 一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现以下步骤：

[0021] 获取待处理的网络通信相关数据；

[0022] 从所述网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；所述网络攻击数据包括在所述网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据；

[0023] 根据所述网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征；

[0024] 对所述网络攻击特征进行分类处理，输出与所述网络攻击事件对应的组织信息；所述组织信息用于对所述网络攻击事件进行追踪溯源。

[0025] 上述网络攻击识别方法、装置、计算机设备和存储介质，从待处理的网络通信相关数据中筛选出与网络攻击事件相关的网络攻击数据，进而根据网络攻击数据确定与相应网络攻击事件对应的网络攻击特征，并对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息。其中，筛选出的网络攻击数据中包括有在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据，这样就可将一系列网络攻击行为相关的多个特征进行了有效关联，在进行组织识别时，可以考虑多个维度的信息，大大提升了组织识别的精准度，能够对恶意组织(比如恶意团队或恶意家族)保持跟踪，实现对网络攻击的精准的追踪溯源。

附图说明

[0026] 图1为一个实施例中网络攻击识别方法的应用环境图；

[0027] 图2为一个实施例中网络攻击识别方法的流程示意图；

[0028] 图3为一个实施例中从网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据的步骤的流程示意图；

[0029] 图4为一个实施例中根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征的步骤的流程示意图；

[0030] 图5为一个实施例中决策树模型的结构示意图；

[0031] 图6为一个实施例中通过决策树进行组织分类的结构示意图；

[0032] 图7(A)为一个具体的实施例中网络攻击识别方法的流程示意图；

[0033] 图7(B)为另一个具体实施例中网络攻击识别方法的流程示意图；

[0034] 图8为一个实施例中网络攻击识别装置的结构框图；

[0035] 图9为另一个实施例中网络攻击识别装置的结构框图；

[0036] 图10为一个实施例中计算机设备的内部结构图。

具体实施方式

[0037] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0038] 本申请提供的网络攻击识别方法，可以应用于如图1所示的应用环境中。其中，终端110通过网络与服务器120进行通信。本申请各实施例所提供的网络攻击识别方法可通过终端110或服务器120单独执行，还可以通过终端110 和服务器120共同协作执行。其中，服务器可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。终端可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表等，但并不局限于此。终端以及服务器可以通过有线或无线通信方式进行直接或间接地连接，本申请在此不做限制。

[0039] 需要说明的是，在一些具体的应用场景中，当服务器为云服务器时，本申请可应用于云计算平台，以为用户提供安全服务。那么在这种情况下，很显然，本申请会涉及到云安全技术。云安全(Cloud Security) 是指基于云计算商业模式应用的安全软件、硬件、用户、机构、安全云平台的总称。云安全融合了并行处理、网格计算、未知病毒行为判断等新兴技术和概念，通过网状的大量客户端对网络中软件行为的异常监测，获取互联网中木马、恶意程序的最新信息，并发送到服务端进行自动分析和处理，再把病毒和木马的解决方案分发到每一个客户端。

[0040] 云安全主要研究方向包括:1. 云计算安全，主要研究如何保障云自身及云上各种应用的安全，包括云计算机系统安全、用户数据的安全存储与隔离、用户接入认证、信息传输安全、网络攻击防护、合规审计等;2. 安全基础设施的云化，主要研究如何采用云计算新建与整合安全基础设施资源，优化安全防护机制，包括通过云计算技术构建超大规模安全事件、信息采集与处理平台，实现对海量信息的采集与关联分析，提升全网安全事件把控能力及风险控制能力;3. 云安全服务，主要研究各种基于云计算平台为用户提供的安全服务，如防病毒服务等。

[0041] 可以理解，在另一些具体的应用场景中，服务器还可以是物理服务器，通过物理服务器为用户提供网络防护安全功能。

[0042] 在一个实施例中，如图2所示，提供了一种网络攻击识别方法，以该方法应用于计算机设备(比如图1中的终端110或服务器120)为例进行说明，该网络攻击识别方法包括以下步骤：

[0043] 步骤S202，获取待处理的网络通信相关数据。

[0044] 其中，网络通信相关数据是设备在进行数据通信的过程中所产生的相关数据，包括通信数据和日志数据。其中，通信数据可以是不同设备间的通信数据，也可以是同一设备内不同程序或进程间的通信数据等，本申请实施例对此不做限定。通信数据具体可包括通信过程中产生的网络流量数据，还可包括有主机设备运行时产生的进程数据或线程数据等。其中，网络流量数据具体可以是通过交换机转发的数据包。进程数据包括进程名、以及进程执行后得到的进程结果等。线程数据包括线程名、以及线程执行后得到的线程结果等。

日志数据是对主机设备的运行状态进行记录的数据，包括硬件状态日志和应用系统日志。硬件状态日志包括主机设备的CPU(central processing unit，中央处理器)或内存使用状态等；应用系统日志包括操作系统、以及应用程序在运行时所产生的日志数据等。

[0045] 具体地，网络设备在运行时，会产生网络通信相关数据，网络设备可采集预设时间内所产生的网络通信相关数据，并将网络通信相关数据传输至计算机设备以进行处理。其中，网络设备具体可以是交换机、路由器、或主机设备等中的至少一种。

[0046] 在一个实施例中，步骤S202，也就是获取待处理的网络通信相关数据的步骤，具体包括：确定部署于预设区域内的网络设备；网络设备包括交换机和主机设备；获取通过网络交换机进行转发的网络流量数据；获取主机设备在运行时产生的进程数据、线程数据和日志数据；将网络流量数据、进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据。

[0047] 可以理解，对于某个企业或者集团，通常会在预设区域内部署一台或多台交换机和主机设备。这些交换机和主机设备可处于同一局域网中，也可以处于广域网中，本申请实施例对此不做限定。

[0048] 其中，部署于预设区域内的交换机用于对不同设备间的通信数据进行转发。具体地，计算机设备可导出交换机中流量层信息，得到对应的网络流量数据。导出的网络流量数据中包括用于通信的数据包。可以理解，当该预设区域中还部署有路由器时，计算机设备也可从路由器中导出对应的网络流量数据。

[0049] 计算机设备还可获取主机设备上的进程数据、线程数据和日志数据。进而，计算机设备将获取的网络流量数据、进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据。

[0050] 在一个实施例中，网络设备可按预设频率定期将该时间周期内产生的网络通信相关数据上报至计算机设备，或者计算机设备可按预设频率定期从网络设备处拉取该时间周期内产生的网络通信相关数据。这样，就可按预设频率定期对网络通信相关数据进行分析，判断是否发生网络攻击事件，并确定发起网络攻击事件的发起方的组织信息，以周期性的对网络攻击进行追踪溯源。

[0051] 上述实施例中，从交换机和主机设备处可获取大量的待处理的网络通信相关数据，以对待处理的网络通信相关数据进行网络攻击分析，可以实现对预设区域内的网络安全进行监控和防护。

[0052] 步骤S204，从网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；网络攻击数据包括在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据。

[0053] 其中，网络攻击事件是由一系列相关的网络攻击行为所构成的事件，这一系列相关的网络攻击行为往往在战术上具有先后关系。比如某次网络攻击事件中就包括有不同阶段的网络攻击行为，如Initial Access(初始访问)阶段中的External Remote Services(外部远程访问)、Privilege Escalation(特权提升)阶段中的Accessibility Features(辅助功能)、以及Collect(数据采集)阶段中的Input Capture(输入获取)等网络攻击行为。这些不同阶段的网络攻击行为具有上下文关系，共同构成一次完整的网络攻击事件。也就是，对于一次完整的网络攻击事件而言，其中发生有至少一种的网络攻击行为。可以理解，网络攻击行为是恶意的网络行为，会对系统带来安全隐患，是需要被识别和防范的。

[0054] 具体地，计算机设备可对网络通信相关数据进行分析，从中筛选出具有攻击行为特点的攻击行为数据。计算机设备可直接根据一系列网络攻击行为所对应的攻击行为数据，作为与网络攻击事件相关的网络攻击数据。计算机设备还可将根据一系列网络攻击行为所对应的攻击行为数据、以及发起这一系列网络攻击行为的发起方所对应的基础设施数据，确定与网络攻击事件相关的网络攻击数据。其中，基础设施数据是发起方所在设备对应的设备信息，具体可以是发起方设备的设备标识，比如发起方设备的序列号、MAC(Media Access Control Address，介质访问控制)地址、或网络地址等。

[0055] 步骤S206，根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征。

[0056] 具体地，计算机设备可将网络攻击数据按照预设置的特征维度进行整理，确定与各个特征维度相对应的数据。进而将与各个特征维度相对应的数据按预设转换规则转换成对应的特征值。计算机设备可按预先设置的各特征维度的顺序，拼接多个特征维度的特征值，得到网络攻击特征。可以理解，网络攻击特征实质上是一组向量。

[0057] 在一个实施例中，计算机设备可预先创建特征值映射表，该特征值映射表中记载有不同特征维度的数据与相应特征值间的对应关系。这样，计算机设备可将网络攻击数据中与各个特征维度相对应的数据按特征值映射表则转换成对应的特征值。

[0058] 在一个实施例中，网络攻击数据包括攻击行为数据和基础设施数据。计算机设备可根据威胁建模模型，将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征。按预设映射规则，将网络攻击数据中的基础设施数据，转换成对应的基础设施特征。再拼接攻击行为特征和基础设施特征，得到与网络攻击事件对应的网络攻击特征。其中，关于威胁建模模型的相关内容在后面的实施例中会进行详细描述。

[0059] 步骤S208，对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息；组织信息用于对网络攻击事件进行追踪溯源。

[0060] 具体地，计算机设备可通过检测模型对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息。可以理解，组织信息也就是发起该次网络攻击事件的群体的家族信息。在网络攻击溯源中，同种家族的攻击行为特征相对于不同家族行为特征具有更高的相似性，因而可通过家族信息来对网络攻击事件进行追踪溯源。其中，检测模型是一种具有分类功能的机器学习模型，具体可以是支持向量机模型、决策树模型、或神经网络模型等，本申请实施例对此不做限定。

[0061] 在一个实施例中，该检测模型可通过训练数据集训练得到。该检测模型的训练步骤包括：获取训练数据集，训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据；根据样本网络攻击数据，确定与样本组织信息对应的样本网络攻击特征；通过样本网络攻击特征和对应的样本组织信息进行模型训练，直至满足训练停止条件时停止训练，得到训练好的检测模型。

[0062] 在一个实施例中，计算机设备可预先收集一些样本数据，比如威胁情报库中家族及其对应的攻击行为数据，以及一些公开的威胁情报文章等。进而，计算机设备可对这些样本数据进行预处理，从中提取出与不同家族分别对应的样本攻击行为数据、以及样本基础设施特征数据，以构建对应的训练数据集。可以理解，此处所提到的家族指的是本申请实施例中所提到的组织。同一家族的网络攻击行为具有更高的相似性和关联性。

[0063] 进而，计算机设备可根据样本网络攻击数据，确定与样本组织信息对应的样本网

络攻击特征，再通过样本网络攻击特征和对应的样本组织信息进行模型训练，直至满足训练停止条件时停止训练，得到训练好的检测模型。

[0064] 其中，训练停止条件是停止模型训练的条件，具体可以是达到预设迭代次数或训练得到的检测模型的分类性能达到预设指标等。

[0065] 这样，通过样本网络攻击特征和对应的样本组织信息进行模型训练，可以训练得到具有组织分类能力的检测模型。在一个实施例，计算机设备在通过样本网络攻击特征和对应的样本组织信息训练得到检测模型后，该检测模型即可用于对网络攻击特征进行分类，以对网络攻击事件进行追踪溯源。

[0066] 可以理解，该检测模型的训练和使用，可以是在同一台计算机设备上执行，也可以是在不同的计算机设备上执行，本申请实施例对此不做限定。

[0067] 在一个实施例中，通过计算机设备输出的组织信息，可以帮助安全人员对网络攻击事件进行快速溯源，识别威胁情况，并且可将结果反馈到检测模型用以对检测模型进行持续更新。

[0068] 上述网络攻击识别方法，从待处理的网络通信相关数据中筛选出与网络攻击事件相关的网络攻击数据，进而根据网络攻击数据确定与相应网络攻击事件对应的网络攻击特征，并通过检测模型对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息。其中，筛选出的网络攻击数据中包括在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据，这样就可将一系列网络攻击行为相关的多个特征进行了有效关联，在进行组织识别时，可以考虑多个维度的信息，大大提升了组织识别的精准度，能够对恶意组织(比如恶意团队或恶意家族)保持跟踪，实现对网络攻击的精准的追踪溯源。

[0069] 在一个实施例中，该网络攻击识别方法具体包括步骤S302至步骤S312，具体如下：

[0070] 步骤S302，获取待处理的网络通信相关数据。

[0071] 步骤S304，确定网络通信相关数据中与不同网络行为分别对应的网络行为数据。

[0072] 其中，网络行为是在网络通信的过程中发生的具体操作，比如访问操作、应用程序启动运行操作、数据发送操作、或数据爬取操作等。可以理解，网络行为包括正常的通信操作，也可包括异常的网络攻击行为。与网络行为对应的网络行为数据具体可以是执行网络行为所产生的数据。比如，当网络操作为访问操作时，对应的网络行为数据可包括有访问页面、访问时间、以及访问频次等；当网络操作为应用程序启动运行操作时，对应的网络行为数据可包括有程序名称、程序来源、以及程序运行状态等数据。

[0073] 具体地，计算机设备可对网络通信相关数据进行分析，确定与不同网络行为分别对应的网络行为数据。也就是以单次的网络行为为单位，将网络通信相关数据划分成多份网络行为数据，这样就可基于每次网络行为所对应的网络行为数据进行专门分析，以从大量的网络通信相关数据中定位到攻击行为数据。

[0074] 在一个实施例中，在步骤S306之前，该方法还包括攻击行为数据的确定步骤，该步骤具体包括：对网络行为数据进行分析，确定各种网络行为分别发生的频次、以及各种网络行为所对应的行为关键信息；当频次大于等于阈值，或者行为关键信息中包括有恶意关键词时，确定相应的网络行为为网络攻击行为；将与网络攻击行为相关的网络行为数据作为攻击行为数据。

[0075] 在一个实施例中，计算机设备可对网络行为数据进行分析，统计各种网络行为分

别发生的频次、以及各种网络行为所对应的行为关键信息。当频次大于等于阈值时，则可确定相应的网络行为为网络攻击行为。当行为关键信息中包括有恶意关键词时，也可确定相应的网络行为为网络攻击行为。进而计算机设备可将与网络攻击行为相关的网络行为数据作为攻击行为数据。

[0076] 在一个实施例中，计算机设备可对网络行为数据进行分析，基于网络攻击行为的行为特征，找到网络行为数据中的攻击行为数据。比如，网络爆破这种攻击行为，具有明显的行为规律，比如在两个相同的网络地址之间会反复执行相同的操作。那么计算机设备可通过统计相同网络行为发生的频次，当频次大于等于阈值时，则判定发生了恶意的网络攻击行为。

[0077] 在一个实施例中，计算机设备可对网络通信相关数据中的日志数据进行分析，确定各网络行为所对应的行为关键信息。行为关键信息比如运行的软件名称或注册表等。当计算机设备从行为关键信息中查找到恶意关键词时，则判定相应的网络攻击行为为网络攻击行为。其中，恶意关键词比如恶意软件名称、或预设的特定词等。

[0078] 在一个实施例中，计算机设备还可对进程数据和线程数据进行分析，当进程名或线程名为恶意攻击行为所对应的进程名或线程名时，则判定相应的网络攻击行为为网络攻击行为。

[0079] 上述实施例中，通过网络行为重复发生的频次，或网络行为数据中是否包括有恶意关键词，可以准确且快速地从大量的网络行为数据中查找出攻击行为数据。

[0080] 步骤S306，当网络行为数据为攻击行为数据时，确定产生攻击行为数据的发起方。

[0081] 具体地，当网络通信相关数据中包括有攻击行为数据时，计算机设备可确定产生该攻击行为数据的发起方，具体可以是确定发起方所对应的设备信息或网址信息等，这些设备信息和网址信息可以标记出发起方。设备信息比如发起方设备的MAC地址；网络信息比如发起方通信时的网络地址或者IP地址所属的网段等。可以理解，发起方，也就是发起网络攻击事件的攻击者，在进行网络攻击时，发起方通常都会持续的发起网络攻击行为。

[0082] 步骤S308，将网络通信相关数据中所有与发起方相关的网络行为数据、以及发起方的基础设施数据，作为与网络攻击事件相关的网络攻击数据。

[0083] 具体地，计算机设备可将网络通信相关数据中所有与该发起方相关的网络行为数据、以及发起方的基础设施数据，作为该发起方所发起的网络攻击事件相关的网络攻击数据。其中，基础设施数据是与发起方所在设备对应的设备信息，具体可以是发起方设备的设备标识，比如发起方设备的序列号、MAC(Media Access Control Address，介质访问控制)地址、或网络地址等。

[0084] 可以理解，在一些实施例中，在一次完整的网络攻击事件中，发起所有网络攻击行为所涉及的基础设施数据是相同的。在另一些实施例中，发起方在发起不同网络攻击行为时所对应的基础设施数据有相同也有不同，本申请实施例对此不做限定。

[0085] 步骤S310，根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征。

[0086] 步骤S312，对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息；组织信息用于对网络攻击事件进行追踪溯源。

[0087] 上述实施例中，将网络通信相关数据中所有与发起网络攻击行为的发起方相关的网络行为数据、以及发起方的基础设施数据，作为与网络攻击事件相关的网络攻击数据，涵

盖了一次网络攻击事件中全方面的相关数据,可以进一步提高组织识别的准确性。

[0088] 在一个实施例中,该网络攻击识别方法具体包括步骤S402至步骤S412,具体如下:

[0089] 步骤S402,获取待处理的网络通信相关数据。

[0090] 步骤S404,从网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据;网络攻击数据包括在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据。

[0091] 步骤S406,获取威胁建模模型,并根据威胁建模模型,将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征。

[0092] 其中,威胁建模模型又可称作ATT&CK(Adversarial Tactics, Techniques, and Common Knowledge)模型,是一个反映各个攻击生命周期的网络攻击行为的模型和知识库。通过该ATT&CK模型,统一了网络攻击行为描述的标准,对攻击行为数据所对应的多种网络攻击行为进行了细分,用以对攻击行为特征进行表示。

[0093] 可以理解,本申请各实施例所提及的网络攻击数据中的攻击行为数据,可以体现TTP中的战术、技术以及其上下文关系。其中,TTP(战术Tactics、技术Techniques和过程Procedures)是描述高级威胁组织及其攻击的重要指标。战术(Tactics)是指威胁行为者试图完成的行为的高层次描述;技术(Techniques)是对导致战术的行为或行动的详细描述;过程(Procedures)是关于威胁执行者如何利用该技术实现其目标的技术细节或指示。

[0094] 当攻击者发起一次网络攻击事件时,可采取不同阶段的战术中的具体的某个技术来实现。那么相应的,对一次网络攻击事件中的多种网络攻击行为所产生的攻击行为数据,进行分析后可转换成对应的TTP特征,也就是本申请各实施例所提及的攻击行为特征。

[0095] 参考表1,表1为一个实施例中ATT&CK模型中对应的战术以及技术特征的示例详情。其中第1行是TTP的战术,每一列是战术中所采用的具体的技术。

[0096] 表1

[0097]

Initial Access	Privilege Escalation	Credential Access	Collection	……
Drive-by Compromise	Access Token Manipulation	Account Manipulation	Audio Capture	
Exploit Public-Facing Application	Accessibility Features	Bash History	Automated Collection	
External Remote Services	AppCert DLLs	Brute Force	Clipboard Data	
……	AppInit DLLs	Forced Authentication	Data from Local System	
	……	Hooking	Data from Removable Media	
		Input Capture	Data Staged	
		Input Prompt	……	
		……		

[0098] 可以理解,表1中仅仅示出了部分战术,比如Initial Access(初始访问)、Privilege Escalation(特权提升)、Credential Access(凭证获取)、以及Collection(数据采集)等战术。每种战术下示意了部分技术,比如Initial Access(初始访问)战术下对应有 Drive-by Compromise(路过式威胁)、Exploit Public-Facing Application(利用公开应用程序)、以及External Remote Services(外部远程服务)等。还比如Privilege Escalation(特权提升)战术下对应有Access Token Manipulation(访问令牌操作)、Accessibility Features(辅助功能)、AppCert DLLs、以及AppInit DLLs等。还比如Credential Access(凭证获取)战术下对应有Account Manipulation(账号操作)、Bash

History(bash历史记录)、Brute Force(暴力破解)、Forced Authentication(强制认证)、Hooking(钩子技术)、Input Capture(输入获取)、以及Input Prompt(输入提示)等技术。Collection(数据采集)战术下对应有Audio Capture(音频抓取)、Automated Collection(自动采集)、Clipboard Data(剪贴板数据)、Data from Local System(采集系统本地数据)、Data from Removable Media(采集可移动媒体数据)、以及Data Staged(数据分段)等技术。

[0099] 可以理解,对于每一次的网络攻击行为所采用的技术,计算机设备均可从ATT&CK模型中找到对应的映射关系。也就是说,每一次网络攻击事件中,发起方均有可能采用某些战术中的某个技术来发起对应的网络攻击行为。比如,一次网络攻击事件中发生的各网络攻击行为在ATT&CK表中映射对应如表2所示,其中字体倾斜且带有下划线的技术为攻击者在发起网络攻击事件时在各个战术中所选择的对应技术。

[0100] 表2

[0101]

Initial Access	Privilege Escalation	Credential Access	Collection	……
Drive-by Compromise	Access Token Manipulation	Account Manipulation	Audio Capture	
Exploit Public-Facing Application	*Accessibility Features*	Bash History	Automated Collection	
External Remote Services	AppCert DLLs	*Brute Force*	Clipboard Data	
……	AppInit DLLs	Forced Authentication	Data from Local System	
	……	Hooking	Data from Removable Media	
		Input Capture	*Input Capture*	
		Input Prompt	……	
		……		

[0102] 并且,ATT&CK模型将每种映射的技术均设置了对应的编号,对应的编号也可理解成该技术所对应的特征值。如在Initial access的战术中,其映射的各项技术对应的编号如表3所示。

[0103] 表3

[0104]

ID	Name
T1189	Drive-by Compromise(路过式威胁)
T1190	Exploit Public-Facing Application(利用公开应用程序)
T1133	External Remote Services(外部远程服务)
T1200	Hardware Additions(硬件利用)
T1091	Replication Through Removable Media(通过可移动媒体复制)
T1193	Spearphishing Attachment(鱼叉式网络钓鱼)
T1192	Spearphishing Link(鱼叉式链接)
T1194	Spearphishing via Service(鱼叉式服务)
T1195	Supply Chain Compromise(供应链攻击)
T1199	Trusted Relationship(信任关系)
T1078	Valid Accounts(有效账号)

[0105] 那么,计算机设备就可将一次网络攻击事件,基于ATT&CK模型,并根据技术编号,将其抽象为一条TTP特征链,进而转换成向量形式的攻击行为特征。比如,将某次网络攻击

事件中的各网络攻击数据转换成对应的TTP特征链:T1133,T1086,T1067,T1182,T1090,T1110,T1135,T1076,T1056,T1483,T1002,T1486。

[0106] 可以理解,当计算机设备将网络攻击数据中的攻击行为数据映射为威胁建模模型中对应的不同战术中的技术时,可根据各技术分别对应的编号确定相应的特征值。可以理解,此处不同战术即对应了不同的特征维度。那么,将不同维度所对应的特征值进行拼接,即可得到相应的攻击行为特征。

[0107] S408,将网络攻击数据中的基础设施数据,转换成对应的基础设施特征。

[0108] 具体地,计算机设备可预先建立基础设施数据与基础设施特征间的映射关系,在需要时,可根据相应的映射关系,确定与基础设施数据对应的基础设施特征。

[0109] 在一个实施例中,对于一次网络攻击事件的发起方,在每次发起网络攻击行为时均使用相同的发起方设备,那么在这种情况下,网络攻击数据中,对应各网络攻击行为的基础设施数据是相同的基础设施数据。计算机设备可直接将该基础设施数据转换成对应的基础设施特征。

[0110] 在一个实施例中,对于一次网络攻击事件的发起方,在每次发起网络攻击行为时可使用不同的发起方设备,或者采用不同的网络地址,那么在这种情况下,网络攻击数据中,对应各网络攻击行为的基础设施数据是不同的基础设施数据。计算机设备可将与各网络攻击行为对应的基础设施数据,分别转换成对应的基础设施特征。

[0111] 在一个实施例中,计算机设备可确定与各个战术中使用技术对应涉及的基础设施数据,并将相应的技术设施数据转换成对应的特征值,如果没有以缺失值代表进行填充。如果同一个技术对应有多个基础设施数据的,则拆分为多条数据,分别构建该特征维度下的特征值。这样,将各个特征维度下的特征值拼接成向量构成基础设施特征。

[0112] S410,根据攻击行为特征和基础设施特征,确定与网络攻击事件对应的网络攻击特征。

[0113] 具体地,计算机设备可根据攻击行为特征和基础设施特征对向量进行填充,构成与网络攻击事件对应的网络攻击特征。

[0114] 在一个实施例中,计算机设备可采用自动化脚本的方式,将网络攻击数据转换成对应的网络攻击特征。其中,每条网络攻击特征均可采用以下方式生成:首先将攻击行为数据与ATT&CK中各项战术中具体的技术相对应,确定当次网络攻击事件中已采用的技术,并确定各技术分别对应的特征值;对于不存在的战术则用缺失值代表进行填充。确定与各个战术中使用技术对应涉及的基础设施数据,并将相应的技术设施数据转换成对应的特征值,如果没有以缺失值代表进行填充。如果同一个技术对应有多个基础设施数据的,则拆分为多条数据,分别构建该特征维度下的特征值。这样,一次网络攻击事件所采用的技术所对应的特征值、以及相应技术所涉及的基础设施所对应的特征值,按顺序拼接成向量构成网络攻击特征。

[0115] 步骤S412,对网络攻击特征进行分类处理,输出与网络攻击事件对应的组织信息;组织信息用于对所述网络攻击事件进行追踪溯源。

[0116] 上述实施例中,通过威胁建模模型可将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征,将网络攻击数据中的基础设施数据,转换成对应的基础设施特征。进而可拼接攻击行为特征和基础设施特征,得到与网络攻击事件对应的网络攻击特征。这样构

建的网络攻击特征，融合了多个维度的特征，使得后续基于网络攻击特征进行组织识别更为准确。

[0117] 在一个实施例中，步骤S208，也就是对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息的步骤，具体包括：获取预先构建的决策树模型；依据网络攻击特征中不同特征维度各自对应的特征值，从决策树模型的根节点开始，不断地自上向下从决策树模型中查找与网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止；将目标叶子节点中存储的组织信息，作为网络攻击事件的发起方所对应的组织信息并输出。

[0118] 其中，决策树模型包括根节点、内部节点和叶子节点。根节点和内部节点用于对输入的网络攻击特征进行分支判断。叶子节点用于存储组织信息。具体地，计算机设备可获取构建好的决策树模型，进而将网络攻击特征输入至决策树模型中，从决策树模型的根节点开始遍历，根据网络攻击特征的指定特征维度所对应的特征值与决策树模型中相应特征维度所对应节点的特征值进行比较，根据比较结果选择选择子树分支(也就是选择相匹配的目标内部节点)，继续进行迭代，直到到达叶子节点后停止。可以理解，到达的叶子节点也就是目标叶子节点，该目标叶子节点中存储的组织信息，就是网络攻击事件的发起方所对应的组织信息。

[0119] 在一个实施例中，依据网络攻击特征中不同特征维度各自对应的特征值，从决策树模型的根节点开始，不断地自上向下从决策树模型中查找与网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止，包括：从决策树模型的根节点开始，根据网络攻击特征中与根节点所对应特征维度的特征值，确定下一层的目标内部节点；根据网络攻击特征中与下一层的目标内部节点所对应特征维度的特征值，确定再下一层的目标内部节点，并不断往下查找与网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止。

[0120] 在一个实施例中，计算机设备通过决策树模型对网络攻击特征进行分类处理时，从决策树模型的根节点开始，根据网络攻击特征中与根节点所对应特征维度的特征值，来确定接下来的分支路径，也就是确定下一层的目标内部节点。然后再根据网络攻击特征中与下一层的目标内部节点所对应特征维度的特征值，确定后面的分支路径，这样逐层进行分支判断，直至达到目标叶子节点时为止。到达的目标叶子节点中存储的组织信息，就是网络攻击事件对应的组织信息。

[0121] 这样，不断的根据网络攻击特征中不同特征维度所对应的特征值，从决策树模型中选择相应的分支路径，可以对网络攻击特征进行准确的分类，得到网络攻击事件的发起方所属组织的组织信息。

[0122] 参考图5，图5为一个实施例中决策树模型的结构示意图。如图5所示，该决策树模型中包括有根节点a、内部节点b-h、以及叶子节点L_1-L_9。可以理解，上述决策树模型的结构、内部节点和叶子节点的数量仅用于进行示意说明，在不同的场景中该决策树模型可以具有更复杂或更简单的分支结构、具有更多数量或更少数量的内部节点和叶子节点，本申请实施例对此不做限定。

[0123] 其中，节点a-h分别表示不同的特征维度，相应的不同分支上的a_1、a_2、a_3、b_1、b_2.....g_1、及h_1分别表示不同的特征值。当某个网络攻击事件对应的网络攻击特征为$a_2b_1c_1d_3e_1f_2g_2h_2$时，那么相应的，计算机设备可从决策树的根节点开始，将与根节点对应特

征维度的特征值，与根节点所存储的各特征值进行比较，选择对应的分支路径。参考图6，图6为一个实施例中通过决策树进行组织分类的结构示意图。如图6所示，根节点对应特征维度a，网络攻击特征为中与特征维度a对应的特征值为a_2，那么相应的，选择与a_2对应分支，也就是走向了内部节点c。进而再将与内部节点c对应特征维度的特征值与内部节点c所存储的各特征值进行比较，选择下面的分支路径。如图6中的选择了与c_1对应分支，这样依次往下不断的选择分支路径，直至达到叶子节点L_7。叶子节点L_7中存储的组织信息就是对应的分类结果。

[0124] 可以理解，图6中分支路径仅为示意性的说明，对于不同的网络攻击特征，决策树模型在进行分类处理时所选中的分支路径也会不同。

[0125] 上述实施例中，通过决策树模型，并根据网络攻击特征中不同特征维度各自对应的特征值，可将网络攻击特征进行快速准确地分类到对应的组织信息上。

[0126] 在一个实施例中，该检测模型具体可以是决策树模，计算机设备可通过以下步骤构建决策树模型：获取训练数据集，训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据；根据样本网络攻击数据，确定与样本组织信息对应的样本网络攻击特征；通过样本网络攻击特征和对应的样本组织信息构建决策树模型。

[0127] 在一个实施例中，样本网络攻击数据包括样本攻击行为数据和样本基础设施数据；根据样本网络攻击数据，确定与样本组织信息对应的样本网络攻击特征，包括：根据威胁建模模型，将样本攻击行为数据映射成对应的样本攻击行为特征；将样本基础设施数据转换成对应的样本基础设施特征；根据样本攻击行为特征和样本基础设施特征，确定与样本组织信息对应的样本网络攻击特征。

[0128] 在一个实施例中，计算机设备可采用与前述实施例中，根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征的相同的方式，将样本网络攻击数据转换成对应的样本网络攻击特征。详细的内容可参考前述实施例中的描述。

[0129] 在一个实施例中，通过样本网络攻击特征和对应的样本组织信息构建决策树模型的步骤，具体包括：确定与样本网络攻击特征对应的多于一个的特征维度；根据训练数据集从特征维度中选择其中一个特征维度作为分类特征以创建根节点，并根据选择的分类特征将训练数据集分裂成多个训练子集；在分裂产生的训练子集中不断的选择分类特征创建内部节点，并根据选择的分类特征进行数据分裂产生新的训练子集，直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上；根据各样本组织信息创建对应的叶子节点；根据创建的根节点、根节点之下的内部节点、以及叶子节点，确定决策树模型。

[0130] 可以理解，在根据训练数据集构建决策树模型，主要就是当一个节点所代表的分类特征无法给出准确的判断时，则选择将这一节点分成多个子节点，这样不断往下划分，直至可以准确的对样本数据进行分类。

[0131] 具体地，计算机设备可先确定与样本网络攻击特征对应的多于一个的特征维度。然后基于训练数据集中带有样本组织信息的样本网络攻击特征，模拟以各个特征维度分别作为分类特征时，对训练数据集进行相应划分后各个训练子集的不确定性。选择不确定性最小的特征维度作为分类特征，并构建根节点。然后基于当前选择的特征维度对训练数据集进行划分，使得训练数据集被分裂成多个训练子集。在分裂过程不断产生的训练子集里重复进行递归，最终完成决策树模型的构建。也就是，对于分裂后的各个训练子集，分别执

行与训练数据集相同的方式，来选择下次分裂的分类特征并进行数据分裂。这样，不断的选择分类特征并进行数据分裂，直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上。

[0132] 可以理解，当某条分支路径无法再分裂时，那么最终到达的节点就是叶子节点，叶子节点中存储有对应的组织信息。计算机设备将根节点、各内部节点和各叶子节点，按照各自对应的父子关系连接起来就构成了决策树模型。

[0133] 在一个实施例中，计算机设备可采取多种方式来确定根据分类标签划分后的训练子集的不确定性。比如，计算机设备可通过训练子集的信息增益、信息增益比或基尼指数等来衡量训练子集的不确定性。当然，计算机设备可采用其他的方式来选择每次分裂的分类特征，本申请实施例对此不做限定。

[0134] 上述实施例中，从根节点开始，基于训练数据集从多个特征维度中选择其中一个特征维度作为分类特征，并根据分类特征将训练数据集分裂成多个训练子集，在分裂不断产生的训练子集里重复递归进行，最终可完成决策树模型的快速而准确的构建。

[0135] 在一个实施例中，根据训练数据集从特征维度中选择其中一个特征维度作为分类特征以创建根节点，包括：根据训练数据集计算每个特征维度分别对应的信息增益率；将信息增益率中的最大信息增益率所对应的特征维度，作为与训练数据集对应的分类特征；根据与训练数据集对应的分类特征创建根节点。

[0136] 在一个实施例中，计算机设备可根据训练数据集计算每个特征维度分别对应的信息增益率，然后将信息增益率中的最大信息增益率所对应的特征维度，作为与训练数据集对应的分类特征，进而根据与该训练数据集对应的分类特征创建根节点。

[0137] 在一个实施例中，对于每个特征维度，计算机设备可通过以下方式计算各个特征维度对应的信息增益率：

[0138] 假设训练数据集为D，某个特征维度为A，用$g(D,A)$表示特征维度A对训练数据集D的信息增益；用 $g_R(D,A)$表示特征维度A对训练数据集D的信息增益率。首先，可通过下面的公式计算训练数据集D的经验熵$Ent(D)$：$Ent(D)=-\sum_{k=1}^{|y|}p_k\,log_2 pk$；其中|y|表示训练数据集中类别的数目；$p_k$表示第k种分类占训练数据集的比例。$Ent(D)$越小，表示D的纯度越高。

[0139] 可以理解，使用某一个特征维度A对训练数据集进行划分后，通常会带来的纯度提高。一般而言，信息增益越大，意味着使用特征维度A来进行划分所获得的“纯度提升”越大。计算机设备可采用以下公式计算信息增益$g(D,A)$：$g(D,A)=Ent(D)-\sum_{v=1}^{V}\frac{|D^v|}{D}Ent(D^v)$；其中，v表示特征维度A中的某个特征值。信息增益= 根节点的信息熵-所有分支节点的信息熵的加权和。加权求和中的系数，也就是 $\frac{|D^v|}{D}$为划分后对应该特征维度中的各特征值所对应的训练子集的样本数量，与划分前的训练数据集中的样本数量的比值。

[0140] 信息增益率 $g_R(D,A)$ =信息增益$g(D,A)$/属性固有值IV(A)。具体地，计算机设备可通过以下公式计算信息增益率：$g_R(D,A)=\frac{g(D,A)}{IV(A)}$；$IV(A)=-\sum_{v=1}^{V}\frac{|D^v|}{D}\log_2\frac{|D^v|}{D}$；其中，特征维度A的可能取值越大，对应的属性固有值IV(A)通常越大，信息增益率偏向于可能取值减少的属性。因而，通过信息增益率来选择合适的分类特征，可构建效果更好的决策树模型。

[0141] 可以理解，在对内部节点进行进一步划分，选择合适的分类特征时，也可采用同样的方式来计算各个特征维度分别对应的信息增益率，选择信息增益率最大的特征维度作为分类特征。

[0142] 上述实施例中，通过信息增益率来选择合适的分类特征，并不断进行分支构建，可使得最终构建得到的决策树模型具有很好的分类效果。

[0143] 在一个实施例中，该网络攻击识别方法还包括更新决策树模型的步骤，该步骤具体包括获取预设时间段内，通过检测模型对网络攻击特征处理所输出的组织信息；根据预设时间段内处理的网络攻击特征和相应输出的组织信息，更新训练数据集，并基于更新后的训练数据集对决策树模型进行更新。

[0144] 具体地，由于家族攻击手法存在变异的可能，计算机设备可在完成检测后将相应的网络攻击特征纳入模型的增量训练，实现决策树模型的持续更新。计算机设备可每隔预设时间段就基于新增的网络攻击特征和相应输出的组织信息更新训练数据集，并基于更新后的训练数据集重新生成新的决策树模型。

[0145] 在一个实施例中，由于训练数据集发生了更新，基于更新后的训练数据集在每次分支选择时，不同特征维度所对应的信息增益率也会发生变化，因而，相应生成的分支路径就会发生变化。

[0146] 可以理解，因为恶意组织为了对抗安全厂商的检测，会持续对攻击手法进行更新，所以对于判断出组织的数据，可采用增量决策树更新的方式，对决策树模型进行持续更新，可以应对某些恶意组织更改攻击上下文中某项战术策略对应技术，实现可持续追踪家族攻击手法变化情况。

[0147] 上述实施例中，通过检测过程持续的数据反馈，能够对恶意组织保持持续的跟踪，实现更精准的网络攻击行为追踪溯源。

[0148] 参考图7(A)，在一个具体实施例中，该网络攻击识别方法具体包括以下步骤：

[0149] S702，获取训练数据集，训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据；样本网络攻击数据包括样本攻击行为数据和样本基础设施数据。

[0150] S704，根据威胁建模模型，将样本攻击行为数据映射成对应的样本攻击行为特征。

[0151] S706，将样本基础设施数据转换成对应的样本基础设施特征。

[0152] S708，根据样本攻击行为特征和样本基础设施特征，确定与样本组织信息对应的样本网络攻击特征。

[0153] S710，确定与样本网络攻击特征对应的多于一个的特征维度，并根据训练数据集计算每个特征维度分别对应的信息增益率。

[0154] S712，将信息增益率中的最大信息增益率所对应的特征维度，作为与训练数据集对应的分类特征。

[0155] S714，根据与训练数据集对应的分类特征创建根节点，并根据选择的分类特征将训练数据集分裂成多个训练子集。

[0156] S716，在分裂产生的训练子集中不断的选择分类特征创建内部节点，并根据选择的分类特征进行数据分裂产生新的训练子集，直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上。

[0157] S718，根据各样本组织信息创建对应的叶子节点，并根据创建的根节点、根节点之下的内部节点、以及叶子节点，确定决策树模型。

[0158] S720，确定部署于预设区域内的网络设备；网络设备包括交换机和主机设备。

[0159] S722，获取通过交换机进行转发的网络流量数据，并获取主机设备在运行时产生的进程数据、线程数据和日志数据。

[0160] S724，将网络流量数据、进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据。

[0161] S726，确定网络通信相关数据中与不同网络行为分别对应的网络行为数据。

[0162] S728，对网络行为数据进行分析，确定各种网络行为分别发生的频次、以及各种网络行为所对应的行为关键信息。

[0163] S730，当频次大于等于阈值，或者行为关键信息中包括有恶意关键词时，确定相应的网络行为为网络攻击行为，并将与网络攻击行为相关的网络行为数据作为攻击行为数据。

[0164] S732，将网络通信相关数据中所有与产生攻击行为数据的发起方相关的网络行为数据、以及发起方对应的基础设施数据，作为与网络攻击事件相关的网络攻击数据。

[0165] S734，获取威胁建模模型，并根据威胁建模模型，将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征。

[0166] S736，将网络攻击数据中的基础设施数据，转换成对应的基础设施特征。

[0167] S738，根据攻击行为特征和基础设施特征，确定与网络攻击事件对应的网络攻击特征。

[0168] S740，将网络攻击特征输入至构建好的决策树模型中，从决策树模型的根节点开始，根据网络攻击特征中与根节点所对应特征维度的特征值，确定下一层的目标内部节点。

[0169] S742，根据网络攻击特征中与下一层的目标内部节点所对应特征维度的特征值，确定再下一层的目标内部节点，并不断往下查找与网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止。

[0170] S744，将目标叶子节点中存储的组织信息，作为网络攻击事件的发起方所对应的组织信息并输出。

[0171] S746，获取预设时间段内，通过检测模型对网络攻击特征处理所输出的组织信息。

[0172] S748，根据预设时间段内处理的网络攻击特征和相应输出的组织信息，更新训练数据集，并基于更新后的训练数据集对决策树模型进行更新。

[0173] 上述网络攻击识别方法，从待处理的网络通信相关数据中筛选出与网络攻击事件相关的网络攻击数据，进而根据网络攻击数据确定与相应网络攻击事件对应的网络攻击特征，并对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息。其中，筛选出的网络攻击数据中包括有在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据，这样就可将一系列网络攻击行为相关的多个特征进行了有效关联，在进行组织识别时，可以考虑多个维度的信息，大大提升了组织识别的精准度，能够对恶意组织(比如恶意团队或恶意家族)保持跟踪，实现对网络攻击的精准的追踪溯源。

[0174] 在一个具体应用场景中，本方案可具体应用于企业内的威胁情报分析，具体可由服务器执行，并通过服务器提供组织信息查询接口，在该具体应用场景中，组织信息具体可

以是家族信息。首先，可将企业交换机处的网络流量数据导入至终端，将企业内各主机设备上的进程数据、线程数据和日志数据等信息也导入至终端。终端将获取的网络流量数据、进程数据、线程数据和日志数据共同作为待检测的网络通信相关数据。终端调用服务器提供的组织信息查询接口，将待检测的网络通信相关数据输入至服务器。服务器调用预先生成的决策树模型，判断攻击行为所属家族，并输出家族信息。根据服务器返回的家族信息，可以帮助安全人员对网络攻击快速溯源，识别威胁情况，并且将结果反馈到决策树模型用以对决策树模型持续更新。

[0175] 其中，关于计算机设备(比如服务器)如何构建决策树模型和使用决策树模型的具体内容，可参考图7(B)，图7(B)为一个具体的实施例中网络攻击识别方法的流程示意图。如图7(B)所示，整个网络攻击识别方法可分为四个部分，包括训练数据集收集、决策树模型的构建、网络攻击识别以及模型更新。下面将从这四个部分来进行详细说明：

[0176] 1、训练数据集收集：

[0177] 计算机设备可预先收集一些现有的样本数据，比如威胁情报库中家族及其对应的攻击行为数据，以及一些公开的威胁情报文章等。这些样本数据具体可以是家族攻击手法上下文，比如下述的a文和b文：

[0178] a．攻击者向受害者邮箱发送带有恶意word文档的email，word文档中包含有精心构造的恶意宏代码，受害者打开word文档并运行宏代码后，主机会主动连接指定的web服务器，下载恶意软件到本地Temp目录下，并强制执行，进行进一步的横向渗透。

[0179] b．攻击者将远控木马程序伪装成“火爆新闻”、“色情内容”等文件名，通过社交网络发送到目标电脑，受害者双击查看文件立刻被安装远控木马。攻击者通过远控木马控制中毒电脑下载挖矿木马，中毒电脑随即沦为矿工。

[0180] 可以理解，当计算机设备收集了这些家族攻击手法上下文后，根据家族攻击手法上下文，将各家族的攻击手法映射为TTP特征，也就是ATT&CK模型中的各个技术。

[0181] 其中，对家族攻击手法所包含TTP特征的提取，可从收集的现有数据中结合专家经验采用自动化脚本的方式输出对应的TTP特征和基础设施特征，以构建对应的样本网络攻击特征。其中生成的每条训练数据(也就是每个样本网络攻击特征)的特征提取规则如下：

[0182] 每条训练数据包括ATT&CK中完整的各个阶段(比如12个阶段)战术过程以及各个阶段对应的基础设施特征；首先从已知的家族攻击手法上下文中提取攻击过程中的攻击行为数据，与ATT&CK中各项技术相对应，对于不存在的战术用预设数值(比如预设的某个缺失值代表)进行填充。并且，提取各个战术中使用技术对应涉及的基础设施数据，如果没有则以缺失值代表进行填充，有多个基础设施数据的则拆分为多条训练数据。这样，每次构成的24维向量作为一条训练数据。

[0183] 2、决策树模型的建立：

[0184] 具体地，计算机设备可根据训练数据集，学习用于识别家族的决策树模型。决策树模型的构建方式如下：

[0185] a．从根节点开始，基于第1部分获取的训练数据集，计算各个TTP特征以及基础设施特征的各个节点的信息增益率；

[0186] b．选取信息增益率最大的节点作为分类特征，通过信息增益率分裂训练数据集为多个训练子集，在分裂不断的产生的训练子集里重复递归进行，最终完成决策树构建。

[0187] 基于先验训练数据集建立的决策树模型，就可作为家族识别器进行家族信息的识别。

[0188] 3、网络攻击识别：

[0189] 首先采集待检测的网络通信相关数据，采用与训练数据集构建的相同方式，构建待检测的网络攻击特征并输入到已生成的决策树模型中进行判定，识别对应的家族。

[0190] 判定过程的具体做法如下：

[0191] a．将网络攻击特征从决策树模型的根节点开始遍历，根据网络攻击特征的指定特征和其特征值与对应决策树节点的特征值进行比较，根据比较结果选择的子树，并按照其值选择输出分支，继续进行迭代；

[0192] b．重复以上过程直到到达叶子节点后停止；

[0193] c．达到叶子节点后即决策树判断完成，叶子节点中存放的就是家族信息，输出叶子节点中的家族信息。

[0194] 4、模型更新：

[0195] 因为恶意家族为了对抗安全厂商的检测，会持续对手法进行更新，所以对于判断出家族的数据可持续收集，以对决策树模型进行反馈更新。对于检测出的网络攻击特征，采用增量决策树更新手法，对决策树模型进行持续更新，可以应对某些恶意家族更改攻击上下文中某项战术策略对应技术，实现可持续追踪家族攻击手法变化情况。

[0196] 本申请各实施例所提供的网络攻击识别方法，是对TTP以及ATT&CK模型落地场景应用，结合了攻击网络攻击的上下文信息进行家族判定，大大提高了网络攻击识别的准确性。

[0197] 应该理解的是，虽然图2-4、以及图7(A)和图7(B)的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图2-4以及图7(A)和图7(B)中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0198] 在一个实施例中，如图8所示，提供了一种网络攻击识别装置800，该装置可以采用软件模块或硬件模块，或者是二者的结合成为计算机设备的一部分，该装置具体包括：获取模块801、筛选模块802、确定模块803和分类模块804，其中：

[0199] 获取模块801，用于获取待处理的网络通信相关数据。

[0200] 筛选模块802，用于从网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；网络攻击数据包括在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据。

[0201] 确定模块803，用于根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征。

[0202] 分类模块804，用于对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息；组织信息用于对网络攻击事件进行追踪溯源。

[0203] 在一个实施例中，获取模块801，具体用于确定部署于预设区域内的网络设备；网

络设备包括交换机和主机设备;获取通过交换机进行转发的网络流量数据;获取主机设备在运行时产生的进程数据、线程数据和日志数据;将网络流量数据、进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据。

[0204] 在一个实施例中,筛选模块802,具体用于确定网络通信相关数据中与不同网络行为分别对应的网络行为数据;当网络行为数据为攻击行为数据时,确定产生攻击行为数据的发起方;将网络通信相关数据中所有与发起方相关的网络行为数据、以及发起方对应的基础设施数据,作为与网络攻击事件相关的网络攻击数据。

[0205] 在一个实施例中,筛选模块802,具体用于对网络行为数据进行分析,确定各种网络行为分别发生的频次、以及各种网络行为所对应的行为关键信息;当频次大于等于阈值,或者行为关键信息中包括有恶意关键词时,确定相应的网络行为为网络攻击行为;将与网络攻击行为相关的网络行为数据作为攻击行为数据。

[0206] 在一个实施例中,确定模块803,具体用于获取威胁建模模型,并根据威胁建模模型,将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征;将网络攻击数据中的基础设施数据,转换成对应的基础设施特征;根据攻击行为特征和基础设施特征,确定与网络攻击事件对应的网络攻击特征。

[0207] 在一个实施例中,分类模块804,具体用于获取预先构建的决策树模型;依据网络攻击特征中不同特征维度各自对应的特征值,从决策树模型的根节点开始,不断地自上向下从决策树模型中查找与网络攻击特征相匹配的目标内部节点,直至达到目标叶子节点时为止;将目标叶子节点中存储的组织信息,作为网络攻击事件的发起方所对应的组织信息并输出。

[0208] 在一个实施例中,分类模块804,具体用于从决策树模型的根节点开始,根据网络攻击特征中与根节点所对应特征维度的特征值,确定下一层的目标内部节点;根据网络攻击特征中与下一层的目标内部节点所对应特征维度的特征值,确定再下一层的目标内部节点,并不断往下查找与网络攻击特征相匹配的目标内部节点,直至达到目标叶子节点时为止。

[0209] 在一个实施例中,该网络攻击识别装置800还包括模型构建模块805,用于获取训练数据集,训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据;根据样本网络攻击数据,确定与样本组织信息对应的样本网络攻击特征;通过样本网络攻击特征和对应的样本组织信息构建决策树模型。

[0210] 在一个实施例中,模型构建模块805,具体用于根据威胁建模模型,将样本攻击行为数据映射成对应的样本攻击行为特征;将样本基础设施数据转换成对应的样本基础设施特征;根据样本攻击行为特征和样本基础设施特征,确定与样本组织信息对应的样本网络攻击特征。

[0211] 在一个实施例中,模型构建模块805,具体用于确定与样本网络攻击特征对应的多于一个的特征维度;根据训练数据集从特征维度中选择其中一个特征维度作为分类特征以创建根节点,并根据选择的分类特征将训练数据集分裂成多个训练子集;在分裂产生的训练子集中不断的选择分类特征创建内部节点,并根据选择的分类特征进行数据分裂产生新的训练子集,直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上;根据各样本组织信息创建对应的叶子节点;根据创建的根节点、根节点之下的内部节点、以及叶

子节点，确定决策树模型。

[0212] 在一个实施例中，模型构建模块805，具体用于根据训练数据集计算每个特征维度分别对应的信息增益率；将信息增益率中的最大信息增益率所对应的特征维度，作为与训练数据集对应的分类特征；根据与训练数据集对应的分类特征创建根节点。

[0213] 参考图9，在一个实施例中，该网络攻击识别装置还包括模型更新模块806，用于获取预设时间段内，通过检测模型对网络攻击特征处理所输出的组织信息；根据预设时间段内处理的网络攻击特征和相应输出的组织信息，更新训练数据集，并基于更新后的训练数据集对决策树模型进行更新。

[0214] 上述网络攻击识别装置，从待处理的网络通信相关数据中筛选出与网络攻击事件相关的网络攻击数据，进而根据网络攻击数据确定与相应网络攻击事件对应的网络攻击特征，并对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息。其中，筛选出的网络攻击数据中包括有在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据，这样就可将一系列网络攻击行为相关的多个特征进行了有效关联，在进行组织识别时，可以考虑多个维度的信息，大大提升了组织识别的精准度，能够对恶意组织(比如恶意团队或恶意家族)保持跟踪，实现对网络攻击的精准的追踪溯源。

[0215] 关于网络攻击识别装置的具体限定可以参见上文中对于网络攻击识别方法的限定，在此不再赘述。上述网络攻击识别装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

[0216] 在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器或终端，其内部结构图可以如图10所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种网络攻击识别方法。

[0217] 本领域技术人员可以理解，图10中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

[0218] 在一个实施例中，还提供了一种计算机设备，包括存储器和处理器，存储器中存储有计算机程序，该处理器执行计算机程序时实现上述各方法实施例中的步骤。

[0219] 在一个实施例中，提供了一种计算机可读存储介质，存储有计算机程序，该计算机程序被处理器执行时实现上述各方法实施例中的步骤。

[0220] 在一个实施例中，提供了一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该计算机设备执行上述各方法实施例中的步骤。

[0221] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以

通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。

[0222] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0223] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

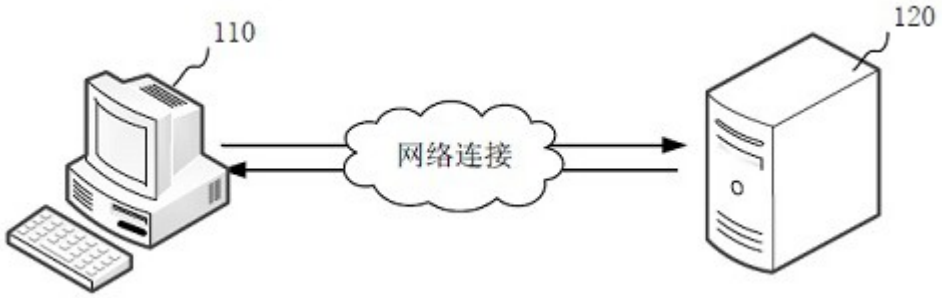

110
120
网络连接

图1

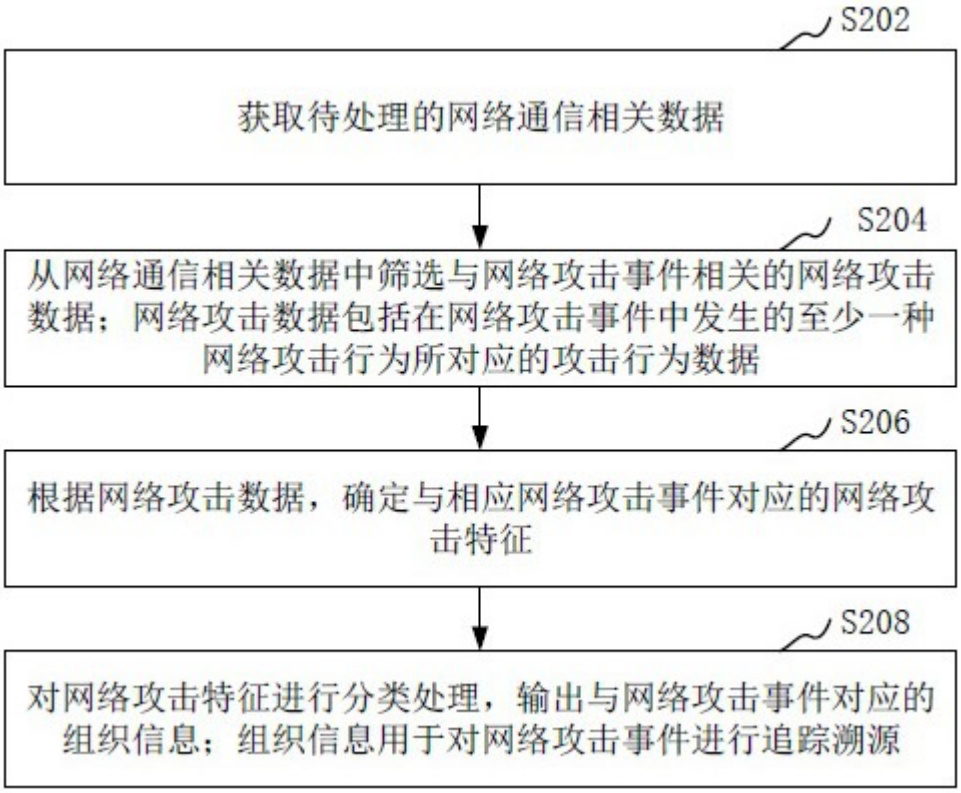

S202
获取待处理的网络通信相关数据
S204
从网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；网络攻击数据包括在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据
S206
根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征
S208
对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息；组织信息用于对网络攻击事件进行追踪溯源

图2

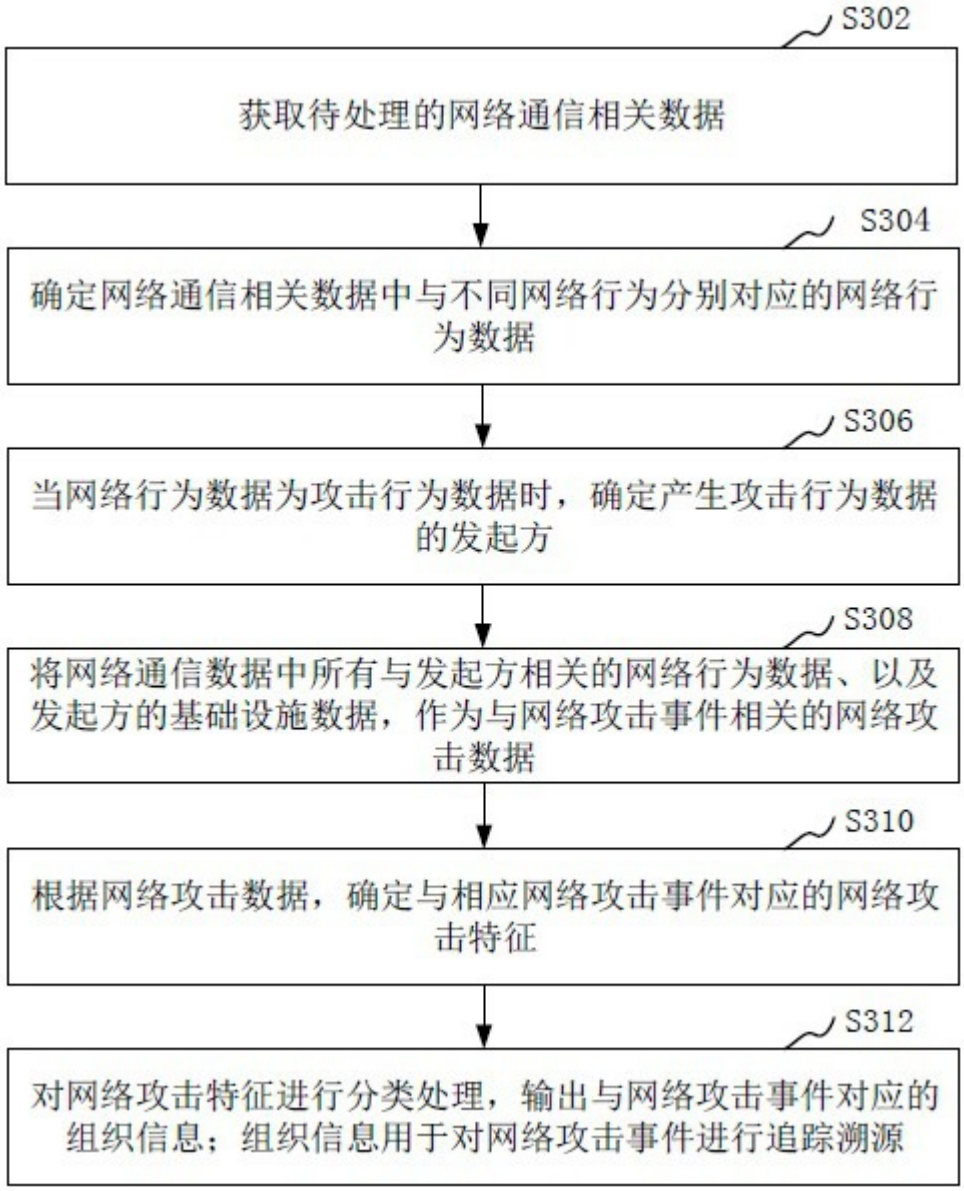

S302
获取待处理的网络通信相关数据
S304
确定网络通信相关数据中与不同网络行为分别对应的网络行为数据
S306
当网络行为数据为攻击行为数据时，确定产生攻击行为数据的发起方
S308
将网络通信数据中所有与发起方相关的网络行为数据、以及发起方的基础设施数据，作为与网络攻击事件相关的网络攻击数据
S310
根据网络攻击数据，确定与相应网络攻击事件对应的网络攻击特征
S312
对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息；组织信息用于对网络攻击事件进行追踪溯源

图3

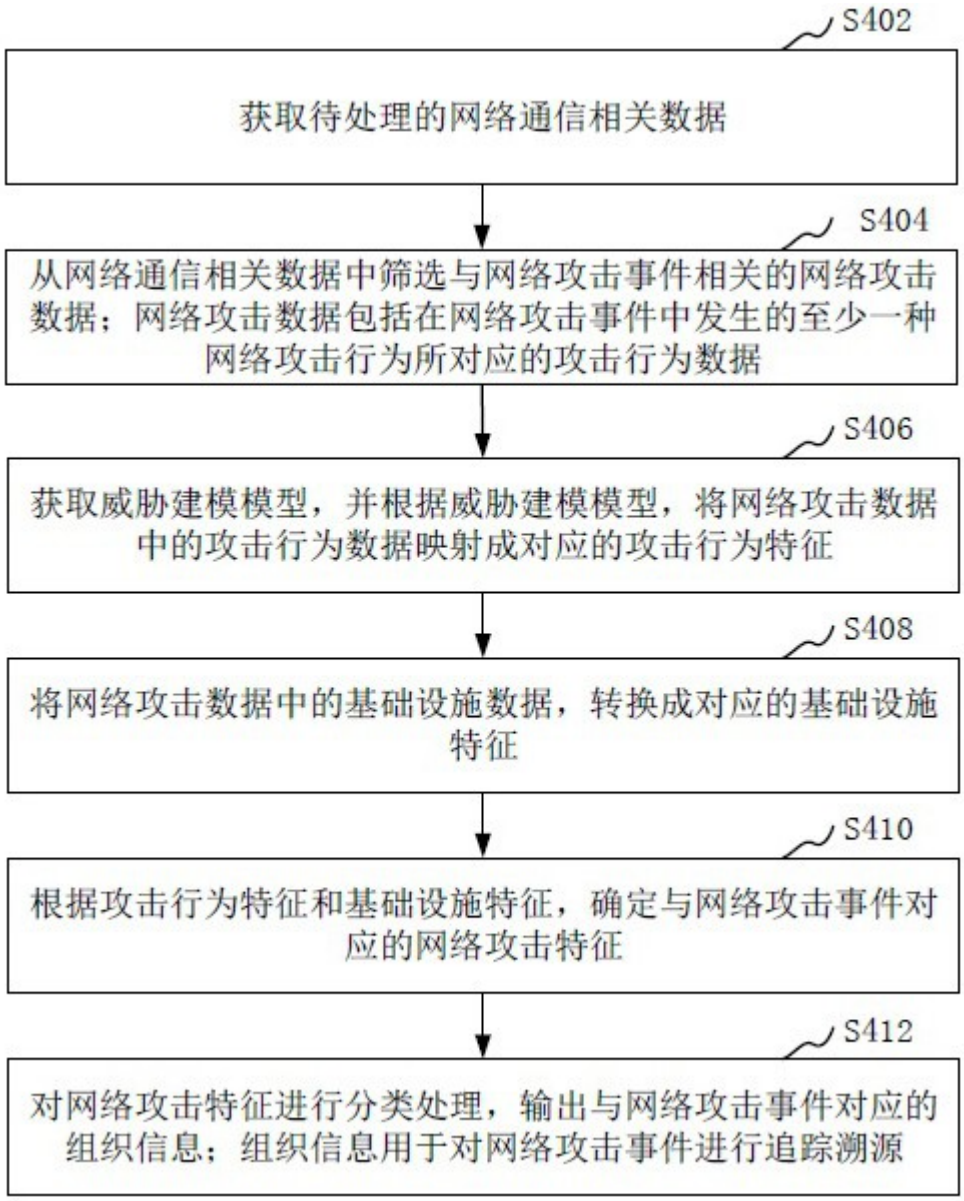

S402
获取待处理的网络通信相关数据
S404
从网络通信相关数据中筛选与网络攻击事件相关的网络攻击数据；网络攻击数据包括在网络攻击事件中发生的至少一种网络攻击行为所对应的攻击行为数据
S406
获取威胁建模模型，并根据威胁建模模型，将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征
S408
将网络攻击数据中的基础设施数据，转换成对应的基础设施特征
S410
根据攻击行为特征和基础设施特征，确定与网络攻击事件对应的网络攻击特征
S412
对网络攻击特征进行分类处理，输出与网络攻击事件对应的组织信息；组织信息用于对网络攻击事件进行追踪溯源

图4

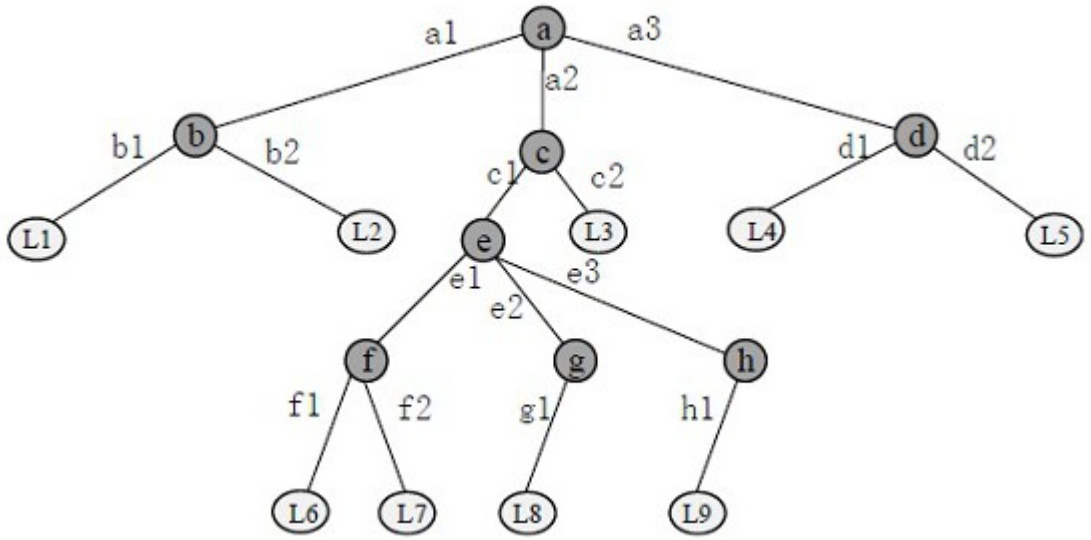

a
a1
a2
a3
b
b1
b2
c
c1
c2
d
d1
d2
L1
L2
e
L3
L4
L5
e1
e2
e3
f
g
h
f1
f2
g1
h1
L6
L7
L8
L9

图5

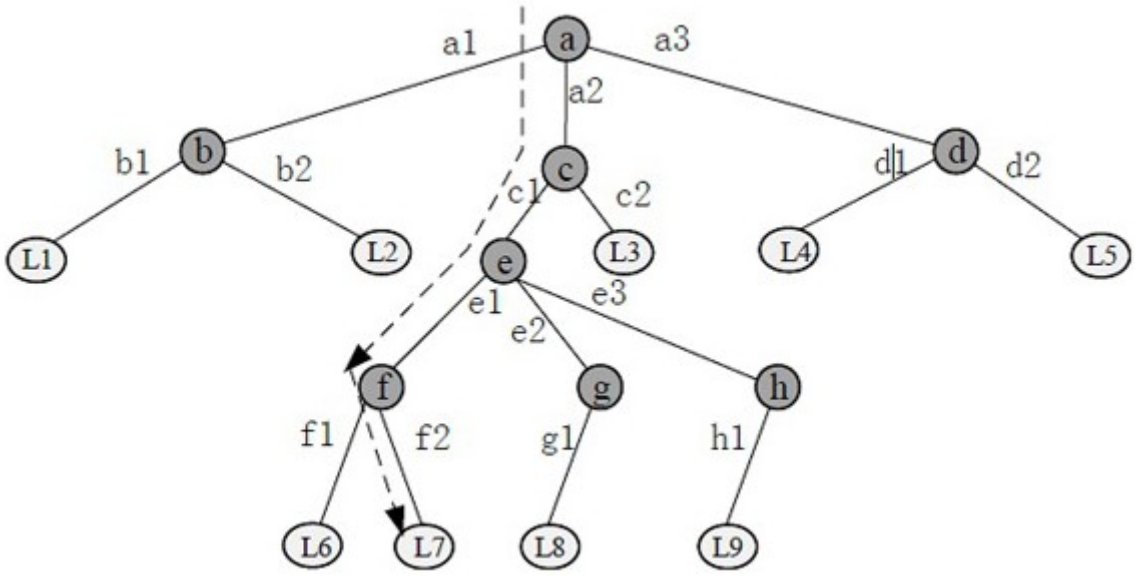

a
a1
a2
a3
b
b1
b2
c
c1
c2
d
d1
d2
L1
L2
e
L3
L4
L5
e1
e2
e3
f
g
h
f1
f2
g1
h1
L6
L7
L8
L9

图6

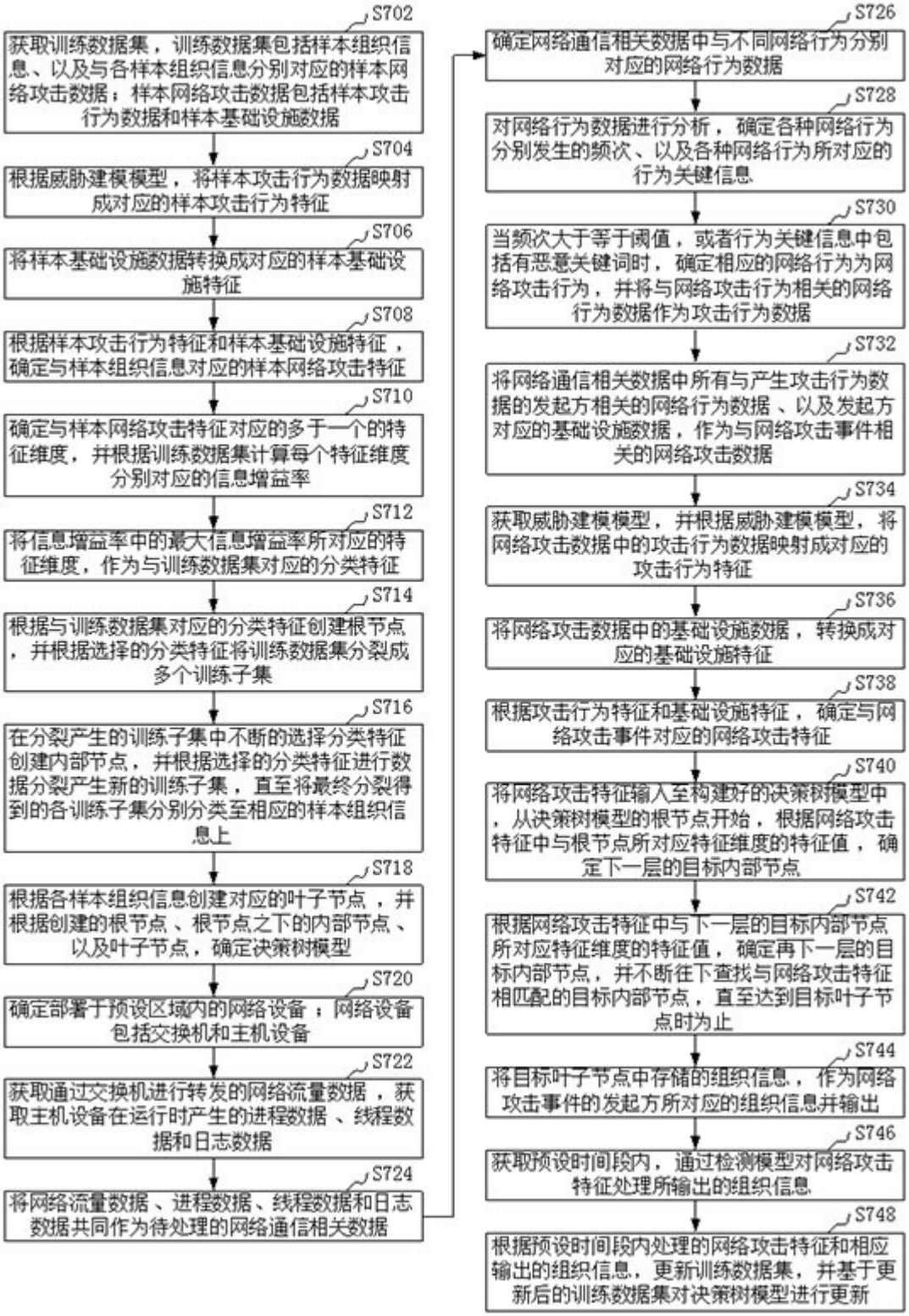

S702
获取训练数据集，训练数据集包括样本组织信息、以及与各样本组织信息分别对应的样本网络攻击数据；样本网络攻击数据包括样本攻击行为数据和样本基础设施数据
S704
根据威胁建模模型，将样本攻击行为数据映射成对应的样本攻击行为特征
S706
将样本基础设施数据转换成对应的样本基础设施特征
S708
根据样本攻击行为特征和样本基础设施特征，确定与样本组织信息对应的样本网络攻击特征
S710
确定与样本网络攻击特征对应的多于一个的特征维度，并根据训练数据集计算每个特征维度分别对应的信息增益率
S712
将信息增益率中的最大信息增益率所对应的特征维度，作为与训练数据集对应的分类特征
S714
根据与训练数据集对应的分类特征创建根节点，并根据选择的分类特征将训练数据集分裂成多个训练子集
S716
在分裂产生的训练子集中不断的选择分类特征创建内部节点，并根据选择的分类特征进行数据分裂产生新的训练子集，直至将最终分裂得到的各训练子集分别分类至相应的样本组织信息上
S718
根据各样本组织信息创建对应的叶子节点，并根据创建的根节点、根节点之下的内部节点、以及叶子节点，确定决策树模型
S720
确定部署于预设区域内的网络设备；网络设备包括交换机和主机设备
S722
获取通过交换机进行转发的网络流量数据，获取主机设备在运行时产生的进程数据、线程数据和日志数据
S724
将网络流量数据、进程数据、线程数据和日志数据共同作为待处理的网络通信相关数据
S726
确定网络通信相关数据中与不同网络行为分别对应的网络行为数据
S728
对网络行为数据进行分析，确定各种网络行为分别发生的频次、以及各种网络行为所对应的行为关键信息
S730
当频次大于等于阈值，或者行为关键信息中包括有恶意关键词时，确定相应的网络行为为网络攻击行为，并将与网络攻击行为相关的网络行为数据作为攻击行为数据
S732
将网络通信相关数据中所有与产生攻击行为数据的发起方相关的网络行为数据、以及发起方对应的基础设施数据，作为与网络攻击事件相关的网络攻击数据
S734
获取威胁建模模型，并根据威胁建模模型，将网络攻击数据中的攻击行为数据映射成对应的攻击行为特征
S736
将网络攻击数据中的基础设施数据，转换成对应的基础设施特征
S738
根据攻击行为特征和基础设施特征，确定与网络攻击事件对应的网络攻击特征
S740
将网络攻击特征输入至构建好的决策树模型中，从决策树模型的根节点开始，根据网络攻击特征中与根节点所对应特征维度的特征值，确定下一层的目标内部节点
S742
根据网络攻击特征中与下一层的目标内部节点所对应特征维度的特征值，确定再下一层的目标内部节点，并不断往下查找与网络攻击特征相匹配的目标内部节点，直至达到目标叶子节点时为止
S744
将目标叶子节点中存储的组织信息，作为网络攻击事件的发起方所对应的组织信息并输出
S746
获取预设时间段内，通过检测模型对网络攻击特征处理所输出的组织信息
S748
根据预设时间段内处理的网络攻击特征和相应输出的组织信息，更新训练数据集，并基于更新后的训练数据集对决策树模型进行更新

图7(A)

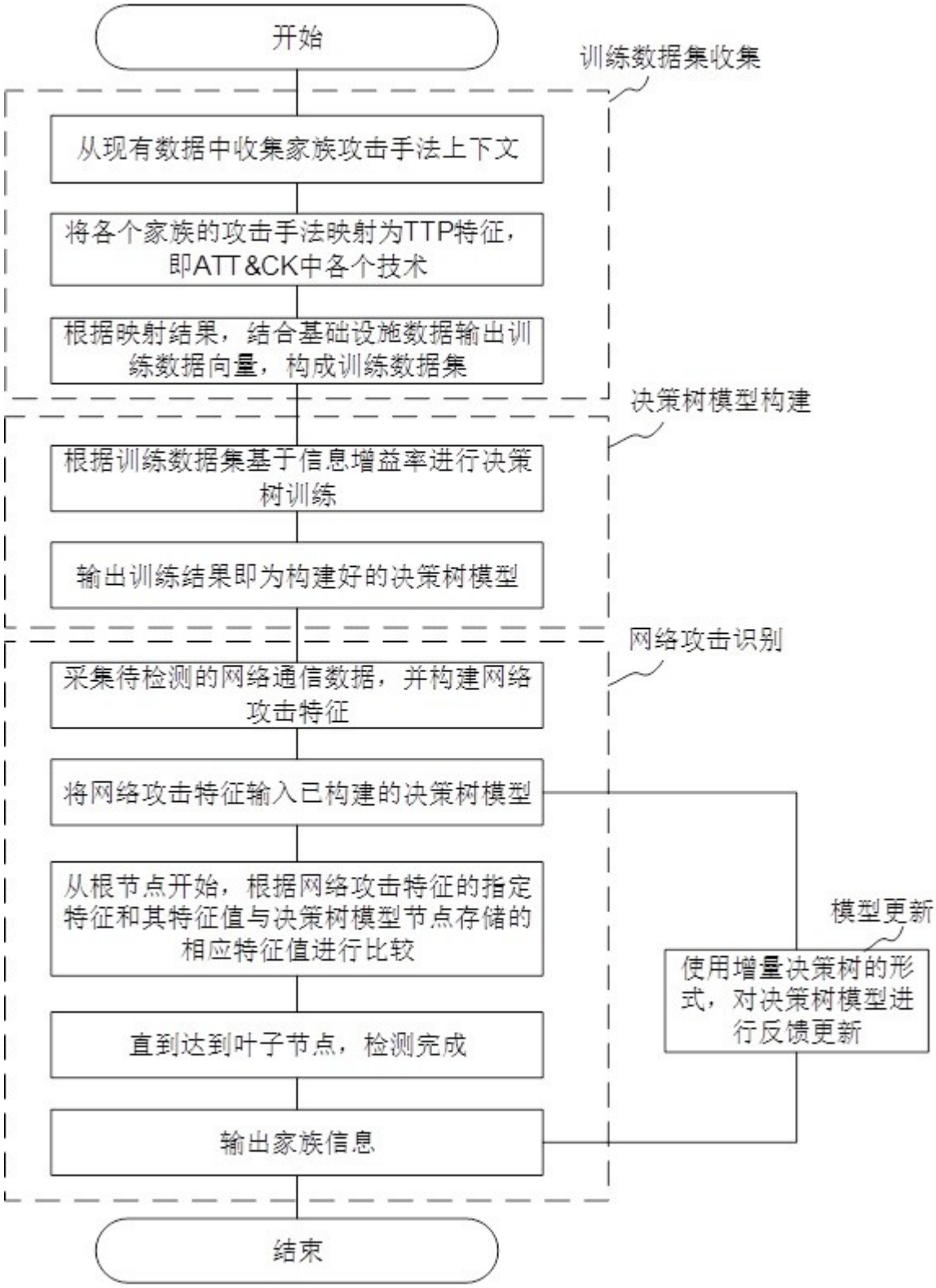

开始
训练数据集收集
从现有数据中收集家族攻击手法上下文
将各个家族的攻击手法映射为TTP特征，即ATT&CK中各个技术
根据映射结果，结合基础设施数据输出训练数据向量，构成训练数据集
决策树模型构建
根据训练数据集基于信息增益率进行决策树训练
输出训练结果即为构建好的决策树模型
网络攻击识别
采集待检测的网络通信数据，并构建网络攻击特征
将网络攻击特征输入已构建的决策树模型
从根节点开始，根据网络攻击特征的指定特征和其特征值与决策树模型节点存储的相应特征值进行比较
直到达到叶子节点，检测完成
输出家族信息
模型更新
使用增量决策树的形式，对决策树模型进行反馈更新
结束

图7(B)

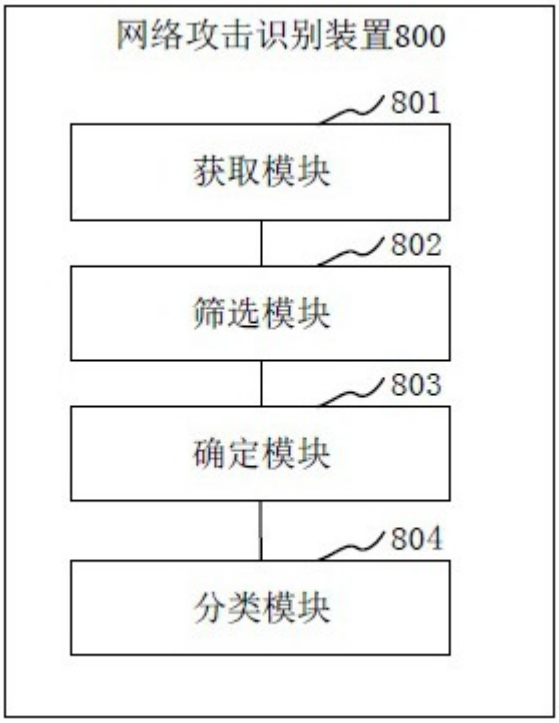

网络攻击识别装置800
801
获取模块
802
筛选模块
803
确定模块
804
分类模块

图8

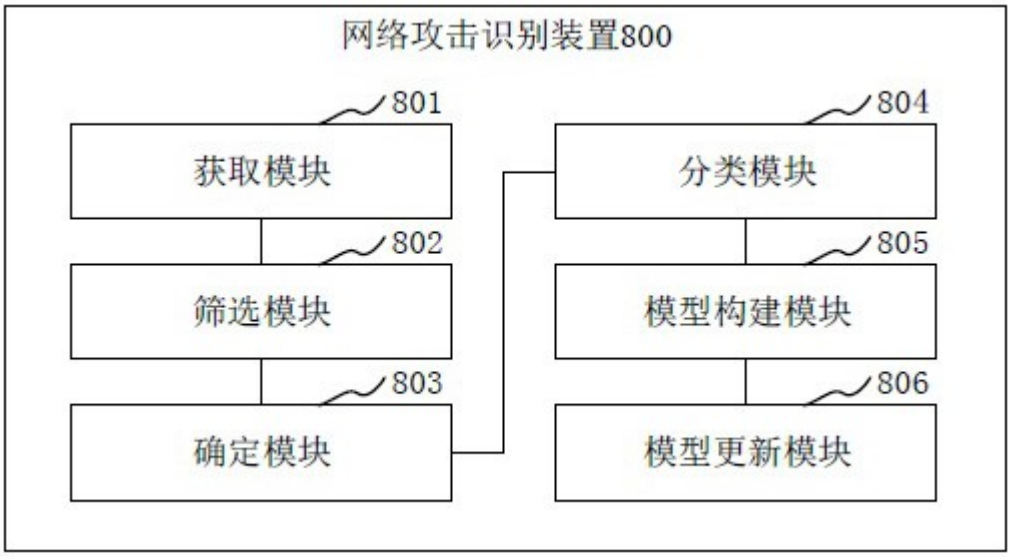

网络攻击识别装置800
801
获取模块
802
筛选模块
803
确定模块
804
分类模块
805
模型构建模块
806
模型更新模块

图9

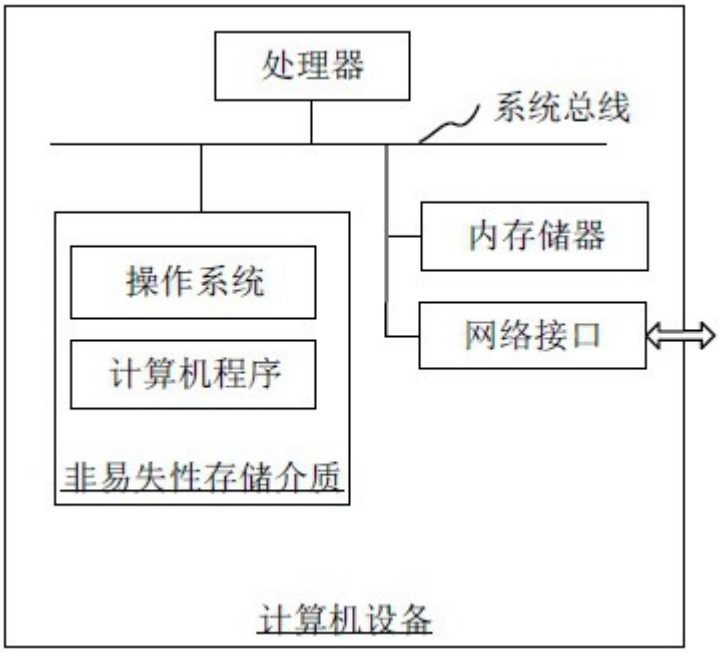


图10