(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115174251 B
(45)授权公告日 2023.09.05

(21)申请号 202210848606.5
(22)申请日 2022.07.19
(65)同一申请的已公布的文献号
申请公布号 CN 115174251 A
(43)申请公布日 2022.10.11
(73)专利权人 深信服科技股份有限公司
地址 518055 广东省深圳市南山区学苑大道1001号南山智园A1栋一层
(72)发明人 姚森友 范炜轩
(74)专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285
专利代理师 王曙聘
(51)Int.Cl.
H04L 9/40(2022.01)
H04L 41/0604(2022.01)
(56)对比文件
CN 114722917 A,2022.07.08
CN 110677433 A,2020.01.10
CN 112738071 A,2021.04.30
CN 110351118 A,2019.10.18
CN 114492653 A,2022.05.13
CN 112613576 A,2021.04.06
CN 110995482 A,2020.04.10
US 2021019372 A1,2021.01.21
WO 2021249629 A1,2021.12.16
US 2020057936 A1,2020.02.20
WO 2021121244 A1,2021.06.24
CN 106254125 A,2016.12.21
CN 112309118 A,2021.02.02
CN 107943856 A,2018.04.20
CN 110309009 A,2019.10.08
US 2017236032 A1,2017.08.17
WO 2021139235 A1,2021.07.15
CN 112596856 A,2021.04.02
CN 106096644 A,2016.11.09
CN 112671767 A,2021.04.16 (续)
审查员 张伟

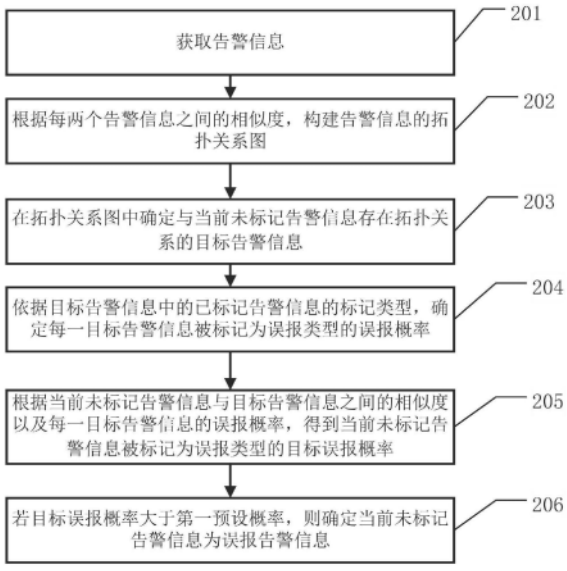
权利要求书2页 说明书10页 附图5页

(54)发明名称

一种安全告警的误报识别方法、装置以及存储介质

(57)摘要

本申请实施例公开了一种安全告警的误报识别方法，用于网络安全技术领域。本申请实施例方法包括：获取告警信息；根据每两个告警信息之间的相似度，构建告警信息的拓扑关系图；在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息；依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为误报类型的误报概率；根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的误报概率，得到当前未标记告警信息被标记为误报类型的目标误报概率；若目标误报概率大于第一预设概率，则确定当前未标记告警信息为误报告警信息，能够有效地从大量的告警信息中识别出误报告警信息。



CN 115174251 B

[转续页]

[接上页]

(56)对比文件

CN 112637178 A,2021.04.09

CN 107451617 A,2017.12.08

CN 112564988 A,2021.03.26

CN 112134799 A,2020.12.25

CN 112333195 A,2021.02.05

CN 109413021 A,2019.03.01

李亚琴;孙传林;雷杰.入侵告警关联系统及关键技术的研究.信息安全与通信保密.2006,(08),第92-94页.

1.一种安全告警的误报识别方法,其特征在于,包括:

获取告警信息,所述告警信息包括第一数量的已标记告警信息和第二数量的未标记告警信息,所述已标记告警信息的标记类型包括误报或非误报;

根据每两个所述告警信息之间的相似度,构建所述告警信息的拓扑关系图;

在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息;其中所述当前未标记告警信息为所述第二数量的未标记告警信息中的任一未标记告警信息;

依据所述目标告警信息中的已标记告警信息的标记类型,确定每一所述目标告警信息被标记为误报类型的误报概率;

根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率,得到所述当前未标记告警信息被标记为误报类型的目标误报概率;

若所述目标误报概率大于第一预设概率,则确定所述当前未标记告警信息为误报告警信息。

2.根据权利要求1所述的误报识别方法,其特征在于,所述获取告警信息包括:

从多个告警日志中提取多个未标记告警信息;

根据预设的研判规则对所述多个未标记告警信息中第一数量的未标记告警信息进行标记,以得到第一数量的已标记告警信息;

将第一数量的所述已标记告警信息和第二数量的所述未标记告警信息确定为所述告警信息;其中所述第一数量与所述第二数量成预设比例关系。

3.根据权利要求1所述的误报识别方法,其特征在于,所述根据每两个告警信息之间的相似度,构建所述告警信息的拓扑关系图包括:

将每个告警信息作为拓扑节点,且为相似度大于预设相似度的两个拓扑节点构建节点连接边,以得到拓扑关系图;

所述在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息包括:

将所述拓扑关系图中与所述当前未标记告警信息存在直接及间接节点连接边的告警信息,确定为所述目标告警信息。

4.根据权利要求1所述的误报识别方法,其特征在于,所述依据所述目标告警信息中的已标记告警信息的标记类型,确定每一所述目标告警信息被标记为误报类型的误报概率包括:

若所述目标告警信息为已标记告警信息,则根据所述目标告警信息本身的标记类型确定所述目标告警信息被标记为误报类型的误报概率;

若所述目标告警信息为未标记告警信息,则在所述拓扑关系图中确定与所述目标告警信息相邻的已标记告警信息,并根据所述相邻的已标记告警信息的标记类型,确定所述目标告警信息被标记为误报类型的误报概率。

5.根据权利要求4所述误报识别方法,其特征在于,所述根据所述相邻的已标记告警信息的标记类型,确定所述目标告警信息被标记为误报类型的误报概率包括:

根据所述相邻的已标记告警信息的标记类型,确定所述相邻的已标记告警信息中误报告警信息的数量占比,并将所述误报告警信息的数量占比作为所述目标告警信息被标记为误报类型的误报概率。

6.根据权利要求1所述的误报识别方法，其特征在于，所述根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率，得到所述当前未标记告警信息被标记为误报类型的目标误报概率包括：

根据与所述当前未标记告警信息相邻的目标告警信息的误报概率，以及所述相邻的目标告警信息与所述当前未标记告警信息之间的相似度，得到所述当前未标记告警信息的初始误报概率；

使用所述当前未标记告警信息所在的目标拓扑关系图中告警信息之间的相似度，每一所述目标告警信息的误报概率，以及所述当前未标记告警信息的初始误报概率，对所述目标拓扑关系图进行更新，直至所述目标拓扑关系图达到收敛时停止更新，得到所述当前未标记告警信息被标记为误报类型的目标误报概率。

7.根据权利要求1所述的误报识别方法，其特征在于，还包括：

若所述当前未标记告警信息为误报告警信息，则将所述当前未标记告警信息存入白名单；

若所述白名单中存在误报告警信息的访问频率达到预设访问条件，则删除所述白名单中访问频率达到所述预设访问条件的误报告警信息。

8.一种安全告警的识别装置，其特征在于，包括：

获取单元，用于获取告警信息，所述告警信息包括第一数量的已标记告警信息和第二数量的未标记告警信息，所述已标记告警信息的标记类型包括误报或非误报；

构建单元，用于根据每两个所述告警信息之间的相似度，构建所述告警信息的拓扑关系图；

第一确定单元，用于在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息；其中所述当前未标记告警信息为所述第二数量的未标记告警信息中的任一未标记告警信息；

第二确定单元，用于依据所述目标告警信息中的已标记告警信息的标记类型，确定每一所述目标告警信息被标记为误报类型的误报概率；

执行单元，用于根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率，得到所述当前未标记告警信息被标记为误报类型的目标误报概率；

第三确定单元，用于若所述目标误报概率大于第一预设概率，则确定所述当前未标记告警信息为误报告警信息。

9.一种安全告警的识别装置，其特征在于，包括：

中央处理器，存储器，输入输出接口，有线或无线网络接口，电源；

所述存储器为短暂存储存储器或持久存储存储器；

所述中央处理器配置为与所述存储器通信，在控制面功能实体上执行所述存储器中的指令操作以执行权利要求1至7中任意一项所述安全告警的识别方法。

10.一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包括指令，当所述指令在计算机上运行时，使得计算机执行如权利要求1至7所述安全告警的识别方法。

一种安全告警的误报识别方法、装置以及存储介质

技术领域

[0001] 本申请实施例涉及网络安全技术领域，尤其涉及一种安全告警的误报识别方法、装置以及存储介质。

背景技术

[0002] 现有的网络安全设备在检测到流量中的攻击行为时会生成安全告警，如代码注入、异常登录、漏洞利用等。网络安全设备会将相应的安全告警信息生成对应的告警日志或告警报告，以便于运维人员等对安全告警进行处理。

[0003] 然而，在复杂的业务场景下，由于网络安全设备对应的客户业务不规范以及部分网络安全设备的检测规则太宽松，不可避免地会在检测过程中产生大量的误报告警，导致真正存在攻击行为的告警被淹没在误报告警中。运营人员面对这大量的安全告警，往往只能研判少量的安全告警数据，无法从告警日志中及时发现威胁，极大影响了运维效率。

[0004] 目前部分厂商对网络安全设备采用源IP、目的IP和攻击类型聚类的方法对安全告警的告警信息进行告警聚合，然而聚类方法的粒度过于粗容易将误报和攻击聚合在一起，若攻击者利用代理不断变换IP对客户端进行攻击时，采用聚类的方法难以有效地从大量的告警信息中识别出误报告警信息。

发明内容

[0005] 本申请实施例提供了一种安全告警的误报识别方法、装置以及存储介质，能够有效地从大量的告警信息中识别出误报告警信息。

[0006] 本申请实施例提供了一种安全告警的误报识别方法，包括：

[0007] 获取告警信息，所述告警信息包括第一数量的已标记告警信息和第二数量的未标记告警信息，所述已标记告警信息的标记类型包括误报或非误报；

[0008] 根据每两个所述告警信息之间的相似度，构建所述告警信息的拓扑关系图；

[0009] 在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息；其中所述当前未标记告警信息为所述第二数量的未标记告警信息中的任一未标记告警信息；

[0010] 依据所述目标告警信息中的已标记告警信息的标记类型，确定每一所述目标告警信息被标记为误报类型的误报概率；

[0011] 根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率，得到所述当前未标记告警信息被标记为误报类型的目标误报概率；

[0012] 若所述目标误报概率大于第一预设概率，则确定所述当前未标记告警信息为误报告警信息。

[0013] 进一步的，所述获取告警信息包括：

[0014] 从多个告警日志中提取多个未标记告警信息；

[0015] 根据预设的研判规则对所述多个未标记告警信息中第一数量的未标记告警信息进行标记，以得到第一数量的已标记告警信息；

[0016] 将第一数量的所述已标记告警信息和第二数量的所述未标记告警信息确定为所述告警信息；其中所述第一数量与所述第二数量成预设比例关系。

[0017] 进一步的，所述根据每两个告警信息之间的相似度，构建所述告警信息的拓扑关系图包括：

[0018] 将每个告警信息作为拓扑节点，且为相似度大于预设相似度的两个拓扑节点构建节点连接边，以得到拓扑关系图；

[0019] 所述在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息包括：

[0020] 将所述拓扑关系图中与所述当前未标记告警信息存在直接及间接节点连接边的告警信息，确定为所述目标告警信息。

[0021] 进一步的，所述依据所述目标告警信息中的已标记告警信息的标记类型，确定每一所述目标告警信息被标记为误报类型的误报概率包括：

[0022] 若所述目标告警信息为已标记告警信息，则根据所述目标告警信息本身的标记类型确定所述目标告警信息被标记为误报类型的误报概率；

[0023] 若所述目标告警信息为未标记告警信息，则在所述拓扑关系图中确定与所述目标告警信息相邻的已标记告警信息，并根据所述相邻的已标记告警信息的标记类型，确定所述目标告警信息被标记为误报类型的误报概率。

[0024] 进一步的，所述根据所述相邻的已标记告警信息的标记类型，确定所述目标告警信息被标记为误报类型的误报概率包括：

[0025] 根据所述相邻的已标记告警信息的标记类型，确定所述相邻的已标记告警信息中误报告警信息的数量占比，并将所述误报告警信息的数量占比作为所述目标告警信息被标记为误报类型的误报概率。

[0026] 进一步的，所述根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率，得到所述当前未标记告警信息被标记为误报类型的目标误报概率包括：

[0027] 根据与所述当前未标记告警信息相邻的目标告警信息的误报概率，以及所述相邻的目标告警信息与所述当前未标记告警信息之间的相似度，得到所述当前未标记告警信息的初始误报概率；

[0028] 使用所述当前未标记告警信息所在的目标拓扑关系图中告警信息之间的相似度，每一所述目标告警信息的误报概率，以及所述当前未标记告警信息的初始误报概率，对所述目标拓扑关系图进行更新，直至所述目标拓扑关系图达到收敛时停止更新，得到所述当前未标记告警信息被标记为误报类型的目标误报概率。

[0029] 进一步的，还包括：

[0030] 若所述当前未标记告警信息为误报告警信息，则将所述当前未标记告警信息存入白名单；

[0031] 若所述白名单中存在误报告警信息的访问频率达到预设访问条件，则删除所述白名单中访问频率达到所述预设访问条件的误报告警信息。

[0032] 本申请实施例还提供了一种安全告警的识别装置,包括:

[0033] 获取单元,用于获取至少一个告警信息,所述告警信息包括第一数量的已标记告警信息和第二数量的未标记告警信息,所述已标记告警信息的标记类型包括误报或非误报;

[0034] 构建单元,用于根据每两个所述告警信息之间的相似度,构建所述告警信息的拓扑关系图;

[0035] 第一确定单元,用于在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息;其中所述当前未标记告警信息为所述第二数量的未标记告警信息中的任一未标记告警信息;

[0036] 第二确定单元,用于依据所述目标告警信息中的已标记告警信息的标记类型,确定每一所述目标告警信息被标记为误报类型的误报概率;

[0037] 执行单元,用于根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率,得到所述当前未标记告警信息被标记为误报类型的目标误报概率;

[0038] 第三确定单元,用于若所述目标误报概率大于第一预设概率,则确定所述当前未标记告警信息为误报告警信息。

[0039] 本申请实施例还提供了一种安全告警的识别装置,包括:

[0040] 中央处理器,存储器,输入输出接口,有线或无线网络接口,电源;

[0041] 所述存储器为短暂存储存储器或持久存储存储器;

[0042] 所述中央处理器配置为与所述存储器通信,在控制面功能实体上执行所述存储器中的指令操作以执行上述安全告警的识别方法。

[0043] 本申请实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质包括指令,当所述指令在计算机上运行时,使得计算机执行上述安全告警的识别方法。

[0044] 从以上技术方案可以看出,本申请实施例具有以下优点:

[0045] 本申请实施例中,获取告警信息;根据每两个告警信息之间的相似度,构建告警信息的拓扑关系图;在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息;依据目标告警信息中的已标记告警信息的标记类型,确定每一目标告警信息被标记为误报类型的误报概率;根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的误报概率,得到当前未标记告警信息被标记为误报类型的目标误报概率;若目标误报概率大于第一预设概率,则确定当前未标记告警信息为误报告警信息。通过告警信息之间的相似度构建拓扑关系图,并在拓扑关系图中使用已标记告警信息确定未标记告警信息的目标误报概率,进而确定未标记告警信息是否为误报告警信息,能够有效地从大量的告警信息中识别出误报告警信息。

附图说明

[0046] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0047] 图1为本申请实施例公开的一个传播告警信息的通信建构图;

[0048] 图2为本申请实施例公开的一个安全告警的误报识别流程图；

[0049] 图3为本申请实施例公开的一个安全告警的非误报识别流程图；

[0050] 图4为本申请实施例公开的一个安全告警的误报删减流程图；

[0051] 图5为本申请实施例公开的一个告警信息的拓扑关系图；

[0052] 图6为本申请实施例公开的一个安全告警的误报识别装置；

[0053] 图7为本申请实施例公开的另一安全告警的误报识别装置。

具体实施方式

[0054] 申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0055] 在本申请实施例的描述中，需要说明的是，术语“中心”、“上”、“下”、“左”、“右”、“竖直”、“水平”、“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系，仅是为了便于描述本申请实施例和简化描述，而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作，因此不能理解为对本申请实施例的限制。

[0056] 在本申请实施例的描述中，需要说明的是，除非另有明确的规定和限定，术语“安装”、“相连”、“连接”应做广义理解，例如，可以是固定连接，也可以是可拆卸连接，或一体地连接；可以是机械连接，也可以是电连接；可以是直接相连，也可以通过中间媒介间接相连，可以是两个元件内部的连通。对于本领域的普通技术人员而言，可以具体情况理解上述术语在本申请实施例中的具体含义。

[0057] 现有的网络安全设备在检测到流量中的攻击行为时会生成安全告警，一般情况下，网络安全设备如：WAF、态势感知、终端木马查杀等设置于客户端中，客户端中出现流量中的攻击行为时网络安全设备会生成安全告警对应的告警信息，对应会存在多个告警信息。然而，多个告警信息中可能存在误报告警信息，为了提高安全运维效率、降低运营成本，需要从多个告警信息中识别出误报告警信息。而一般情况下，对于误报告警信息的识别可以在服务器端也可以在客户端，如图1所示：服务器101可以与多个客户端102连接，该连接可以是有线或无线连接，具体此处不做限定。服务器101可以收集多个客户端102生成的告警信息，然后对收集的告警信息进行识别；或者，多个客户端102之间连接，每个客户端102都可以收集其他客户端102生成的告警信息并对本身的加其他客户端102的告警信息进行识别；又或者，客户端102与服务器101连接，可以从服务器101中获取多个告警信息并对告警信息进行识别；客户端102也可以对本身生成的多个告警信息进行识别。

[0058] 现有技术中，对于多个告警信息的识别一般采用聚类的方法对告警信息进行告警聚合，聚类一般指，根据在数据中发现的描述对象及其关系的信息，将数据对象分组。其目标是，组内的对象相互之间是相似的(相关的)，而不同组中的对象是不同的(不相关的)。组内的相似性(同质性)越大，组间差别越大，聚类就越好。然而，聚类的方法粒度过于粗容易将误报和攻击聚合在一起，采用聚类的方法难以有效地从大量的告警信息中识别出误报告警信息，客户端依旧会存在“误报刷屏”的现象。因此，本申请实施例提供了一种安全告警的误报识别方法，能够有效地从大量的告警信息中识别出误报告警信息，如图2所示，具体步骤如下：

[0059] 201、获取告警信息。

[0060] 本申请实施例的误报识别装置可以为服务器或客户端，具体此处不做限定；误报识别装置可以获取至少一个告警信息，具体的，误报识别装置可以从服务器或客户端中获取网络安全设备检测到流量中的攻击行为时生成的多个告警信息。该告警信息中包括第一数量的已标记告警信息和第二数量的未标记告警信息，该第一数量可以为20个或30个，具体此处不做限定，该第二数量可以为200个或300个，具体此处此处不做限定；该已标记告警信息的标记类型包括误报或非误报。可以理解的是，标记类型为误报的已标记告警信息表示已被确定且标记为误报告警信息，标记类型为非误报的已标记告警信息表示已被确定且标记为非误报告警信息，具体的，可以使用“1”将告警信息标记为误报告警信息，使用“0”将告警信息标记为非误报告警信息，具体的标记方式此处不做限定。该已标记告警信息可以理解为已打标签的告警信息，标签类型即为标记类型。该误报告警信息一般指的是网络安全设备在检测中出现了错误，将流量中的非攻击行为生成安全告警。

[0061] 具体的，误报识别装置可以从多个告警日志(安全日志)中提取多个未标记告警信息；该多个告警日志可以为多个客户端中网络安全设备生成，也可以从服务器中获取预先存储的告警日志，具体此处不做限定；每个告警日志包括有多个未标记告警信息。误报识别装置也可以从一个告警日志中提取多个未标记告警信息。误报识别装置根据预设的研判规则对多个未标记告警信息中第一数量的未标记告警信息进行标记，以得到第一数量的已标记告警信息；可以理解的是，误报识别装置可以根据预设的研判规则确定第一数量的每一未标记告警信息是否为误报，根据确定结果标记相应的标记类型。具体的，可以使用标记模块对第一数量的未标记告警信息进行标记，标记模块包括：强规则模块、人工研判模块以及反馈模块；其中，强规则模块通过大量客户端数据的运营，有些规则在客户端对于告警信息的误报概率极低，这类规则可以定义为强规则。只要当前的告警日志中告警信息对应的是强规则，我们就可以直接研判为攻击，即该告警信息为非误报告警信息，不需要再经过人工判定。人工研判模块利用专家经验进行判定。在运营人员对相关告警信息进行研判后，该人工研判模块会对跟已研判告警信息相似的同类告警信息进行聚合，避免重复判定的同时提高了人工研判的质量和效率。反馈模块为当误报识别装置识别出未标记告警信息是否为误报后，根据识别结果对未标记告警信息标记相应的标记类型，并反馈迭代告警日志，进一步增加标记的数量，从而提高标签传播的有效性。

[0062] 将第一数量的已标记告警信息和第二数量的未标记告警信息确定为至少一个告警信息；其中，该第一数量与该第二数量成预设比例关系。该预设比例关系可以为1:40或1:50，具体此处不做限定。可以理解的是，本申请实施例使用的主要为标签传播算法，标签传播算法是半监督算法，需要一定量的标签数据才能将信息传播下去，否则模型可能无法收敛或有较大偏差。因此，在构建拓扑关系图之前需要计算当前已打标签的告警信息的占比是多少。一个可行的实施例是，当已标记告警信息与未标记告警信息的比例超过1:50(预设比例关系为1:50)且告警信息数量(即第一数量加第二数量)大于1000情况下才进行传播学习，这样能保证标签传播算法正常运行。

[0063] 202、根据每两个告警信息之间的相似度，构建告警信息的拓扑关系图。

[0064] 误报识别装置可以根据每两个告警信息之间的相似度，构建告警信息的拓扑关系图。具体的，可以将每个告警信息作为拓扑节点，该告警信息主要包括源sip、目的ip、攻击

类型、攻击荷载、数据包中各个数据字段的数据，具体此处不做限定；且为相似度大于预设相似度的两个拓扑节点构建节点连接边，以得到拓扑关系图。其中，相似度大小是通过最终在数据集中的表现，选择效果最好的数值，该预设相似度可以为50％或60％，具体此处不做限定。具体的，节点连接边由任意两个节点之间的相似度构成，且保留相似度大于预设相似度的节点连接边作为拓扑关系图中节点之间的权重，迭代当前的所有告警日志构建完全图模型。其中，如图5所示为一个构建完成的拓扑关系图，其中，节点501、节点504、节点506以及节点507为未标记节点(未标记告警信息)，节点502、节点505以及节点508为标记为误报的节点(误报告警信息)，节点503以及节点509为标记为非误报的节点(非误报告警信息)，节点与节点之间的节点连接边存在相应的相似度作为权重值。可以理解的是，构建完成的拓扑关系图可以有多个。

[0065] 203、在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息。

[0066] 误报识别装置可以在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息，其中该当前未标记告警信息为第二数量的未标记告警信息中的任一未标记告警信息，即当前未标记节点为第二数量的未标记节点中的任一未标记节点。

[0067] 具体的，误报识别装置可以将拓扑关系图中与当前未标记告警信息存在直接及间接节点连接边的告警信息，确定为目标告警信息。可以理解的是，该目标告警信息可以为已标记告警信息或未标记告警信息。如图5所示，若当前未标记节点为节点501，则节点502、节点503、节点504以及节点505与节点501存在直接节点连接边，而节点506、节点507、节点508以及节点509与节点501存在间接节点连接边，因此，节点502-509均为目标节点(目标告警信息)。

[0068] 204、依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为误报类型的误报概率。

[0069] 误报识别装置可以依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为误报类型的误报概率。可以理解的是，目标告警信息中可能存在有已标记告警信息以及未标记告警信息，本申请实施例主要使用已标记告警信息的标记类型确定每一目标告警信息被标记为误报类型的误报概率。可以将已标记的节点和未标记的节点按误报或非误报的标记类型进行分类，定义一个(i+u)*C的矩阵，这里的i代表已经标记的标签，u代表未标记的标签，C代表类别，可以计算出当前节点会被标记的类别概率，该类别概率中包括有误报概率。

[0070] 具体的，若目标告警信息为已标记告警信息，则根据目标告警信息本身的标记类型确定目标告警信息被标记为误报类型的误报概率；若目标告警信息为未标记告警信息，则在拓扑关系图中确定与目标告警信息相邻的已标记告警信息，并根据相邻的已标记告警信息的标记类型，确定目标告警信息被标记为误报类型的误报概率。可以理解的是，相邻的已标记告警信息为在拓扑关系图中与目标告警信息存在直接节点连接边的已标记告警信息。若目标告警信息为未标记告警信息，具体可以根据该相邻的已标记告警信息的标记类型，确定相邻的已标记告警信息中误报告警信息的数量占比，并将误报告警信息的数量占比作为目标告警信息被标记为误报类型的误报概率。如图5所示，若当前未标记节点(当前未标记告警信息)为节点501，则目标节点(目标告警信息)中，节点502的标记类型为误报则对应的误报概率为100％，节点503的标记类型为非误报则对应的误报概率为0，节点504中

相邻的已标记节点(相邻的已标记告警信息)为节点508以及节点509,则节点504的误报概率为1/2,其他节点的误报概率类似。

[0071] 205、根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的误报概率,得到当前未标记告警信息被标记为误报类型的目标误报概率。

[0072] 误报识别装置可以根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的误报概率,得到当前未标记告警信息被标记为误报类型的目标误报概率。可以理解的是,在标签传播过程中节点与节点之间的权重值也可以作为节点与节点之间的传播概率,权重越大传播概率就越大。可以在当前未标记告警信息(当前未标记节点)所处的拓扑关系图中,限定已标记节点的误报概率,对每个未标记节点按传播概率,将与未标记节点存在拓扑关系的节点的误报概率按权重值相加,并不断更新,直至拓扑关系图收敛。

[0073] 具体的,误报识别装置可以根据与当前未标记告警信息相邻的目标告警信息的误报概率,以及相邻的目标告警信息与当前未标记告警信息之间的相似度,得到当前未标记告警信息的初始误报概率。一般情况下,将相邻的目标告警信息的误报概率按相似度作为权重值进行加权平均,得到当前未标记告警信息的初始误报概率。如图5所示,若当前未标记节点为节点501,则节点501的初始误报概率为(1*65%+1/2*70%+1*80%+0*70%)/4=45%。接着,可以使用当前未标记告警信息所在的目标拓扑关系图中告警信息之间的相似度,每一目标告警信息的误报概率,以及当前未标记告警信息的初始误报概率,对目标拓扑关系图进行更新,直至目标拓扑关系图达到收敛时停止更新,得到当前未标记告警信息被标记为误报类型的目标误报概率。可以理解的是,构建完成的拓扑关系图可能有多个,当前未标记告警信息所在的目标拓扑关系图表示,在目标拓扑关系图中所有的拓扑节点(节点)都与该当前未标记告警信息所对应的当前未标记节点存在拓扑关系。对目标拓扑关系图进行更新指的是对目标拓扑关系图中所有的未标记节点的误报概率进行不断更新,每次更新时都是将与未标记节点相邻节点的误报概率进行加权平均。目标拓扑关系图达到收敛可以为当前未标记节点的误报概率达到收敛,或与当前未标记节点相邻的误报概率达到收敛,具体此处不做限定。该收敛可以理解为在一定次数的更新后的多次更新中,误报概率在一定范围内浮动即可确定为收敛。

[0074] 206、若目标误报概率大于第一预设概率,则确定当前未标记告警信息为误报告警信息。

[0075] 误报识别装置可以根据当前未标记告警信息的目标误报概率确定当前未标记告警信息是否为误报告警信息。具体的,若目标误报概率大于第一预设概率,则确定当前未标记告警信息为误报告警信息。该第一预设概率可以为80%或85%,具体此处不做限定。

[0076] 本申请实施例中,通过告警信息之间的相似度构建拓扑关系图,并在拓扑关系图中使用已标记告警信息确定未标记告警信息的目标误报概率,进而确定未标记告警信息是否为误报告警信息,能够有效地从大量的告警信息中识别出误报告警信息。在一种可实现的方案中,通过对安全日志进行关联,建立拓扑关系,并利用已有的误报标记节点和临近未标注节点的相关性,识别出哪些是误报数据,给未标记的节点打上误报标签。能在误报标记数据较少的情况下,识别出更多的误报数据,从而降低告警误报,有效提高企业的安全运维效率,降低运营成本。适用于多个场景,包括应用层防火墙、态势感知等。所需的标签数量较

少，算法复杂度也较低，在兼顾人力成本和性能的同时在效果方面也较为显著，能够应对日益变化的攻击和业务场景。

[0077] 上述的申请实施例中提出了如何从多个告警信息的未标记告警信息中识别出误报告警信息，进一步的，本申请实施例还可以从多个告警信息的未标记告警信息中识别出非误报告警信息，如图3所示，具体步骤如下：

[0078] 301、获取告警信息。

[0079] 302、根据每两个告警信息之间的相似度，构建告警信息的拓扑关系图。

[0080] 303、在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息。

[0081] 可以理解的是，步骤301至步骤303与上述步骤201至步骤203类似，具体此处不再赘述。

[0082] 304、依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为非误报类型的非误报概率。

[0083] 误报识别装置可以依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为非误报类型的非误报概率，具体与步骤204类似。若目标告警信息为已标记告警信息，则根据目标告警信息本身的标记类型确定目标告警信息被标记为非误报类型的非误报概率；若目标告警信息为未标记告警信息，则在拓扑关系图中确定与目标告警信息相邻的已标记告警信息，并根据相邻的已标记告警信息的标记类型，确定目标告警信息被标记为非误报类型的非误报概率。如图5所示，若当前未标记节点为节点501，则目标节点中，节点502的非误报概率为0，节点503的非误报概率为100％，节点504的非误报概率为1/2。

[0084] 305、根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的非误报概率，得到当前未标记告警信息被标记为非误报类型的目标非误报概率。

[0085] 误报识别装置根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的非误报概率，得到当前未标记告警信息被标记为非误报类型的目标非误报概率。具体与步骤205类似，具体的，可以在当前未标记节点所在的目标拓扑图中，将与每一未标记节点相邻节点的非误报概率以节点之间的相似度作为权重值进行加权平均，并不断进行加权平均，更新至目标拓扑图收敛，得到目标非误报概率。

[0086] 306、若目标非误报概率大于第二预设概率，则确定当前未标记告警信息为非误报告警信息。

[0087] 误报识别装置可以根据当前未标记告警信息的目标非误报概率确定当前未标记告警信息是否为非误报告警信息。若目标非误报概率大于第二预设概率，则确定当前未标记告警信息为非误报告警信息。该第二预设概率可以为75％或80％，具体此处不做限定。

[0088] 进一步的，当从多个告警信息的未标记告警信息中识别出误报告警信息以及非误报告警信息后，将未标记告警信息根据识别结果存入黑白名单，并使用黑白名单更新告警日志，具体如图4所示：

[0089] 当误报识别装置获取到安全日志后，可以使用标记模块对安全日志中第一数量的未标记告警信息进行研判后标记，在拓扑图构建模块中根据安全日志中每个两个告警信息之间的相似度构建拓扑关系图，标签传播模块基于拓扑关系图中告警信息之间的相似度以及已标记告警信息的标记类型，确定未标记告警信息的标记类型。可以理解的是，一般情况

下,当目标误报概率小于等于第一预设概率时,并不表示该当前未标记告警信息确定为非误报告警信息;当目标非误报概率小于等于第二预设概率时,并不表示该当前未标记告警信息确定为误报告警信息。一般情况下,对当前未标记告警信息进行识别时,会同时判断是否满足目标误报概率大于第一预设概率的条件以及目标非误报概率大于第二预设概率的条件,一般情况下,只会满足其中一种条件。当两个条件都不满足时,则需要更多的已标注告警信息并再次识别当前未标记告警信息。

[0090] 若当前未标记告警信息为误报告警信息,则将当前未标记告警信息标记相应的误报类型并存入白名单;若当前未标记告警信息为非误报告警信息,则将当前未标记告警信息标记相应的非误报类型并存入黑名单。可以理解的是,当标签传播模块更新收敛后,会对当前未标记节点输出相应的黑标签概率(目标非误报概率)以及白标签概率(目标误报概率),优选的,将黑白标签概率大于80%的告警信息存入黑白名单中保存。黑白名单用于标记误报告警信息,一个可行的实施例是通过四元组:(url+目的ip+匹配的规则或引擎sid+攻击字段)来标记误报告警信息。黑白名单可以更新标记模块中的告警信息,增加已标记告警信息的数量,有效提高误报识别的准确性。

[0091] 白名单方便下次研判的时候进行提前过滤,即若有网络安全设备检测出的告警信息与白名单中的误报告警信息一致,则可确定该告警信息为误报告警信息,可以删除掉该误报告警信息。然而,白名单满的时候也有可能删除掉刷屏的某段时间的访问数据,因此需要制定相应的删除策略。若白名单中存在误报告警信息的访问频率达到预设访问条件,则删除白名单中访问频率达到所述预设访问条件的误报告警信息。具体的,可以使用最近最少访问(LRU)的删减策略,LRU是一种常用的页面置换算法,选择最近最久未使用的页面予以淘汰。该算法赋予每个页面一个访问字段,用来记录一个页面自上次被访问以来所经历的时间t,当须淘汰一个页面时,选择现有页面中其t值最大的,即最近最少使用的页面予以淘汰。进一步的,使用LRU的同时考虑访问频率,如果当前要删除的数据访问量远远大于倒数第二个要删除的数据访问量,那么考虑删除倒数第二个。如果当前要删除的数据触发了多次这种删除操作,再删除这条数据。这样能一定程度上缓解删除间断刷屏的数据被移出白名单。

[0092] 本申请实施例中,以在标注数据较少的情况下,通过样本间的关系,建立拓扑关系图,并利用已标记的误报节点标签信息来预测未标记节点的标签信息,当生成黑白名单后可以使用黑白名单删除告警日志中的误报告警信息,最终达到误报消减的效果。

[0093] 本申请实施例还提供了一种安全告警的识别装置,如图6所示,包括:

[0094] 获取单元601,用于获取至少一个告警信息,所述告警信息包括第一数量的已标记告警信息和第二数量的未标记告警信息,所述已标记告警信息的标记类型包括误报或非误报;

[0095] 构建单元602,用于根据每两个所述告警信息之间的相似度,构建所述告警信息的拓扑关系图;

[0096] 第一确定单元603,用于在所述拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息;其中所述当前未标记告警信息为所述第二数量的未标记告警信息中的任一未标记告警信息;

[0097] 第二确定单元604,用于依据所述目标告警信息中的已标记告警信息的标记类型,

确定每一所述目标告警信息被标记为误报类型的误报概率；

[0098] 执行单元605，用于根据所述当前未标记告警信息与所述目标告警信息之间的相似度以及每一所述目标告警信息的误报概率，得到所述当前未标记告警信息被标记为误报类型的目标误报概率；

[0099] 第三确定单元606，用于若所述目标误报概率大于第一预设概率，则确定所述当前未标记告警信息为误报告警信息。

[0100] 本申请实施例还提供了一种安全告警的识别装置700，如图7所示，包括：

[0101] 中央处理器701，存储器702，输入输出接口703，有线或无线网络接口704，电源705；

[0102] 所述存储器702为短暂存储存储器或持久存储存储器；

[0103] 所述中央处理器701配置为与所述存储器702通信，在控制面功能实体上执行所述存储器702中的指令操作以执行上述安全告警的识别方法。

[0104] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0105] 在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0106] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0107] 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0108] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM，read-only memory)、随机存取存储器(RAM，random access memory)、磁碟或者光盘等各种可以存储程序代码的介质。

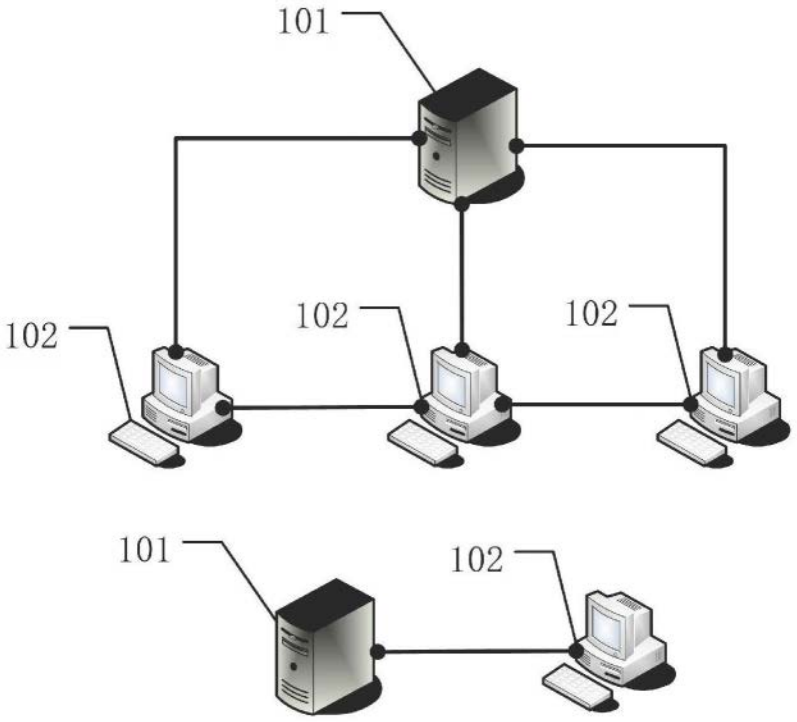


图1

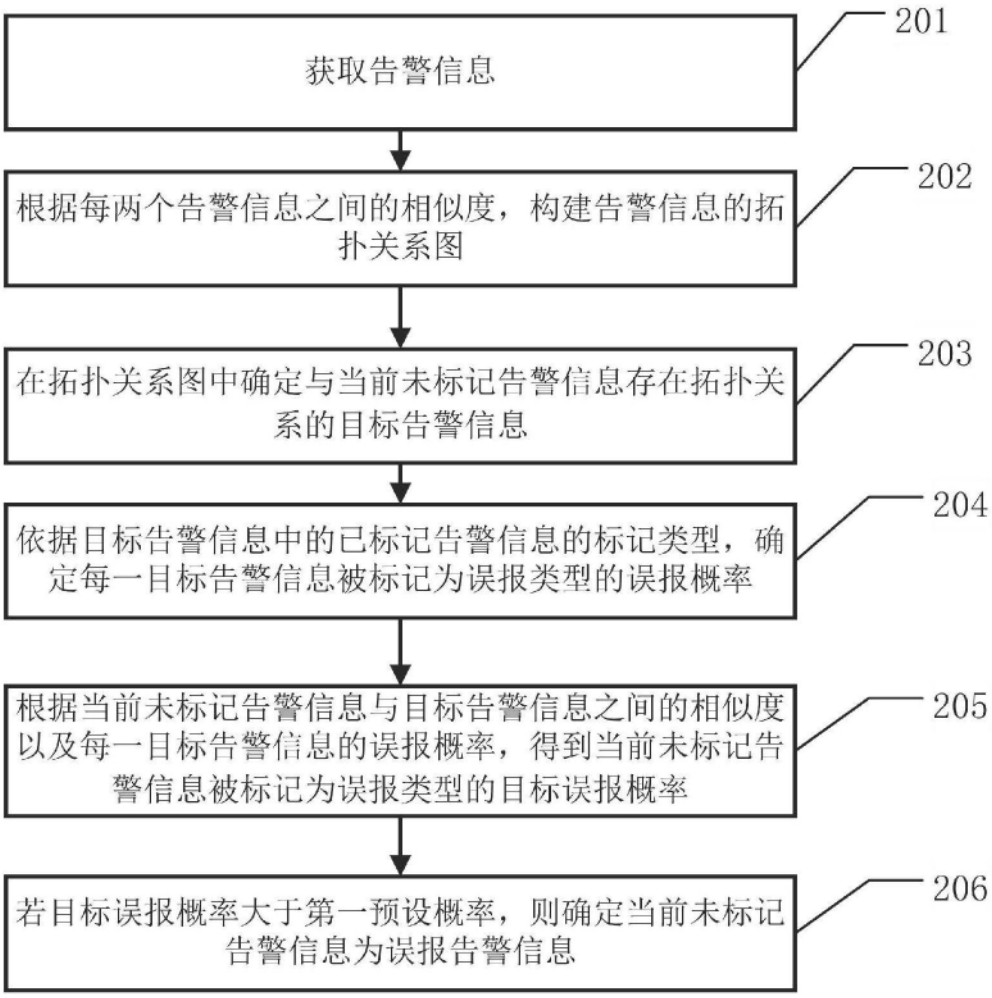

获取告警信息
201
根据每两个告警信息之间的相似度，构建告警信息的拓扑关系图
202
在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息
203
依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为误报类型的误报概率
204
根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的误报概率，得到当前未标记告警信息被标记为误报类型的目标误报概率
205
若目标误报概率大于第一预设概率，则确定当前未标记告警信息为误报告警信息
206

图2

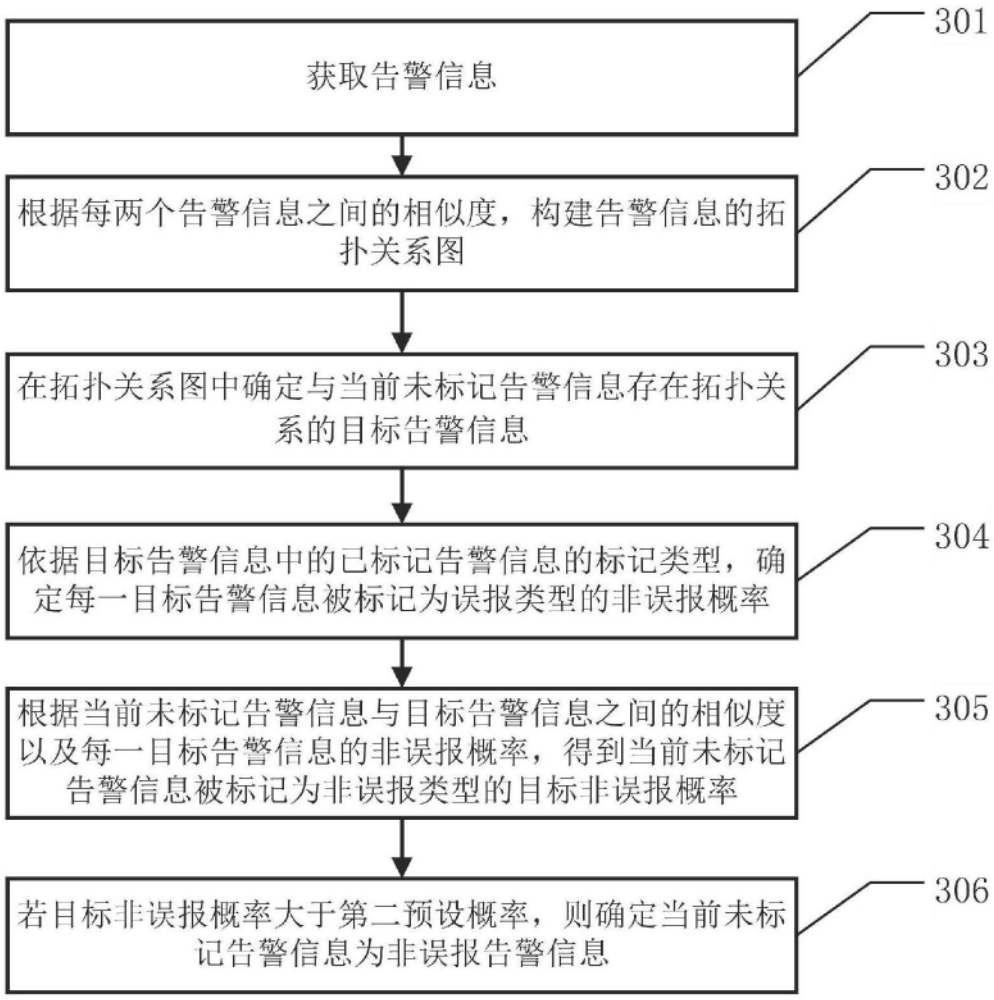

301
获取告警信息
302
根据每两个告警信息之间的相似度，构建告警信息的拓扑关系图
303
在拓扑关系图中确定与当前未标记告警信息存在拓扑关系的目标告警信息
304
依据目标告警信息中的已标记告警信息的标记类型，确定每一目标告警信息被标记为误报类型的非误报概率
305
根据当前未标记告警信息与目标告警信息之间的相似度以及每一目标告警信息的非误报概率，得到当前未标记告警信息被标记为非误报类型的目标非误报概率
306
若目标非误报概率大于第二预设概率，则确定当前未标记告警信息为非误报告警信息

图3

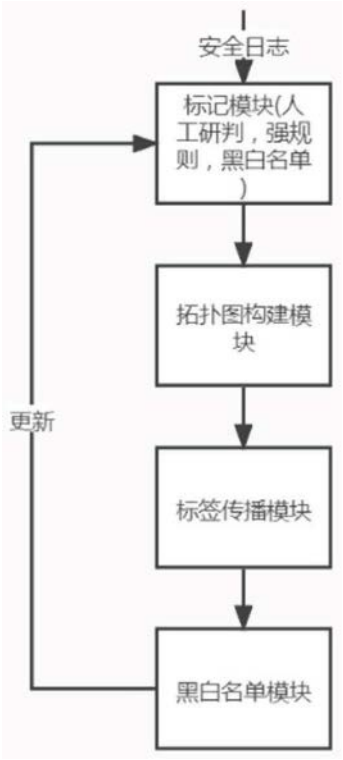

安全日志
标记模块(人工研判，强规则，黑白名单)
拓扑图构建模块
标签传播模块
黑白名单模块
更新

图4

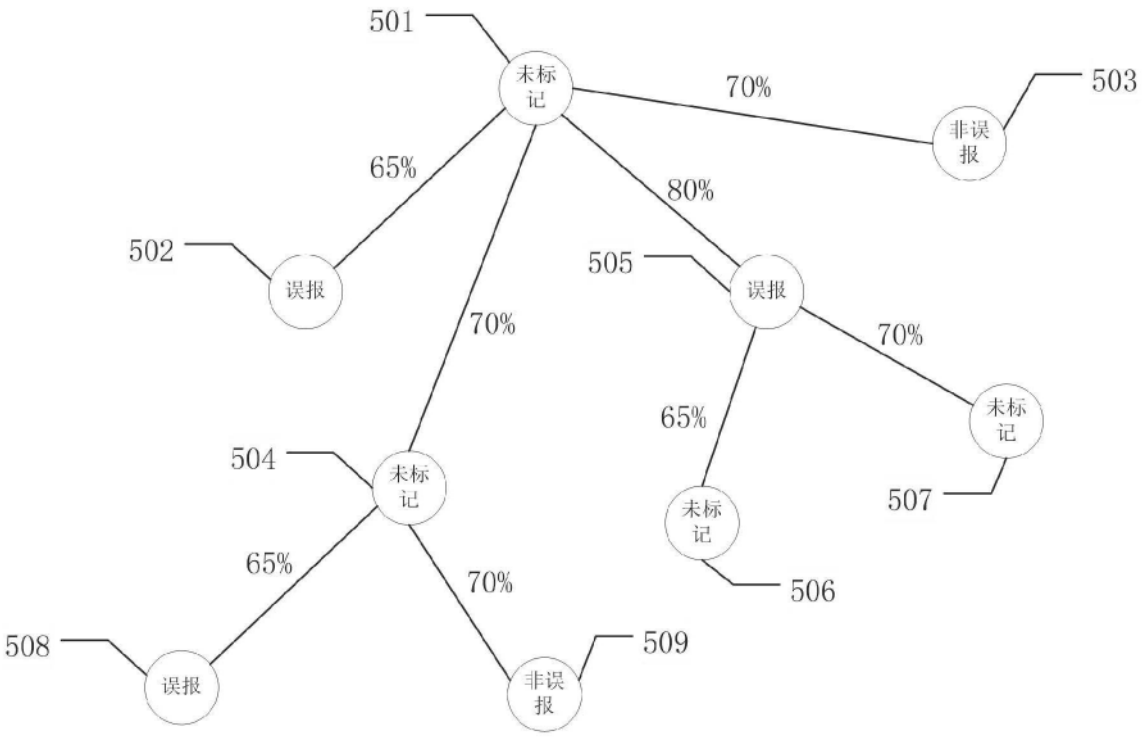

501
未标记
70%
503
非误报
65%
80%
502
误报
505
误报
70%
70%
65%
未标记
504
未标记
507
未标记
506
65%
70%
508
误报
非误报
509

图5

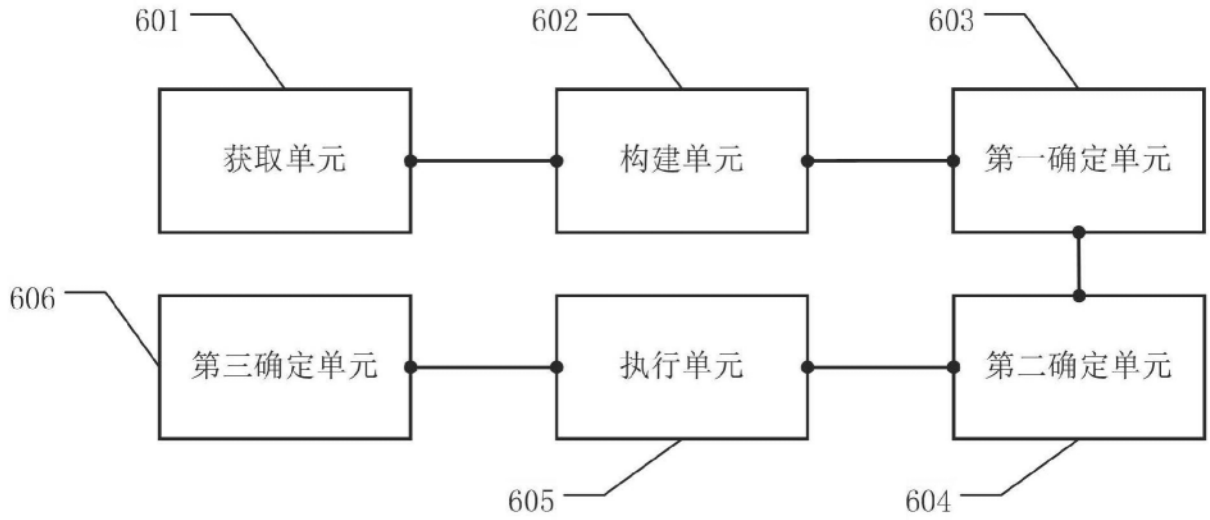

601
602
603
获取单元
构建单元
第一确定单元
606
第三确定单元
执行单元
第二确定单元
605
604

图6

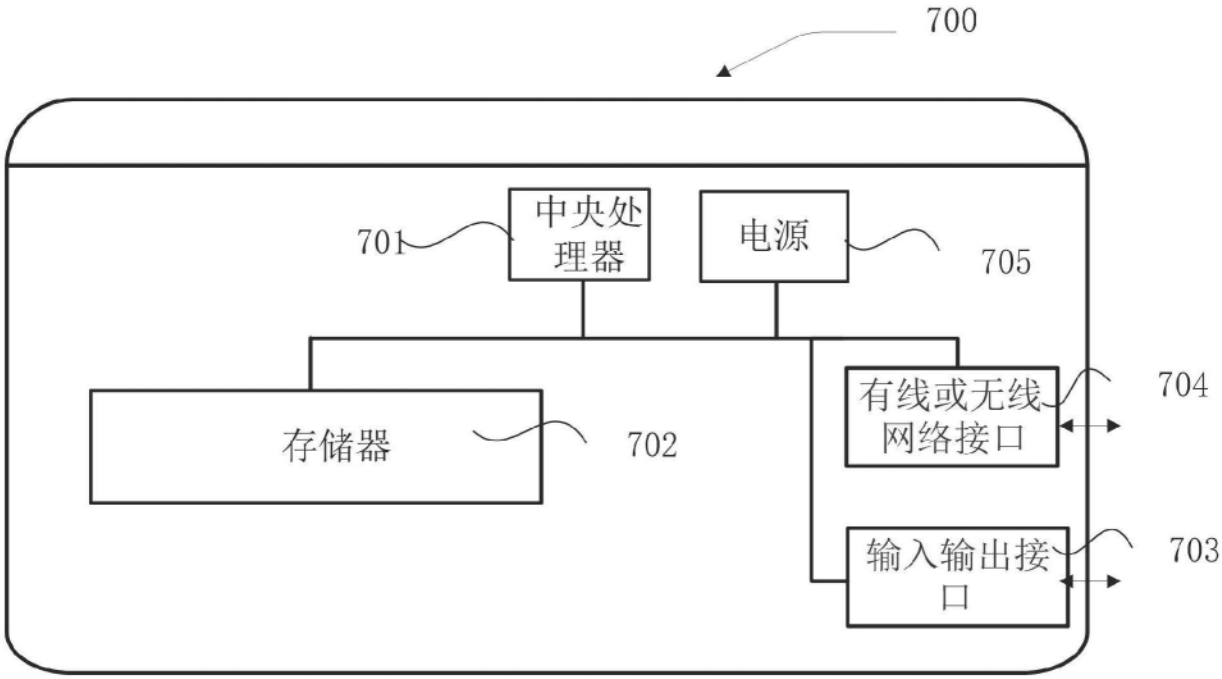

700
701
中央处理器
电源
705
存储器
702
有线或无线网络接口
704
输入输出接口
703

图7