(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109450671 A
(43)申请公布日 2019.03.08

(21)申请号 201811226511.X

(22)申请日 2018.10.22

(71)申请人 北京安信天行科技有限公司
地址 100000 北京市海淀区北四环西路68号10层1001号

(72)发明人 彭海龙 邢亚君 孟铭 吴晓东

(74)专利代理机构 北京高沃律师事务所 11569
代理人 张海青

(51)Int.Cl.
H04L 12/24(2006.01)
H04L 29/06(2006.01)
G06F 16/18(2019.01)

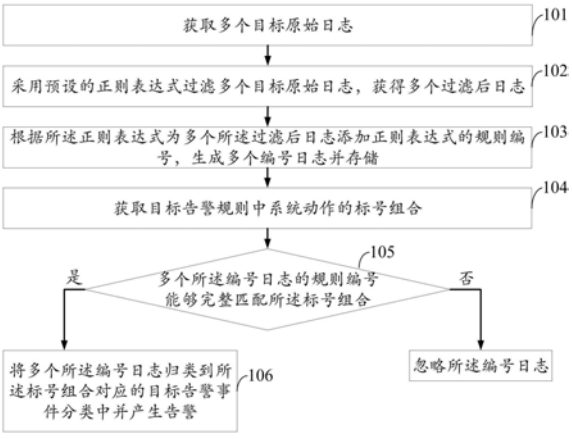
权利要求书3页 说明书13页 附图4页

(54)发明名称

一种日志多组合告警归类方法及系统

(57)摘要

本发明公开了一种日志多组合告警归类方法及系统。所述方法首先获取多个目标原始日志;并采用预设的正则表达式过滤多个目标原始日志,获得多个过滤后日志;根据正则表达式为多个过滤后日志添加正则表达式的规则编号,生成多个编号日志;然后获取目标告警规则中系统动作的标号组合;判断多个编号日志的规则编号是否能够完整匹配所述标号组合,若是,将多个编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。本发明提供的方法通过判断目标日志的规则编号是否能够完整匹配系统动作的标号组合,而不仅仅匹配一个动作,从而避免了现有的日志归类过程中,归类结果准确性不高,且存在误判、漏判的问题,提高了日志归类、告警的准确性。



CN 109450671 A

1.一种日志多组合告警归类方法,其特征在于,所述方法包括:

获取多个目标原始日志;多个所述目标原始日志来源于防火墙、网络设备、主机系统、数据库或中间件;

采用预设的正则表达式过滤多个所述目标原始日志,获得多个过滤后日志;

根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号,生成多个编号日志并存储;

获取目标告警规则中系统动作的标号组合;

判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合,获得第一判断结果;

若所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合,将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。

2.根据权利要求1所述的日志多组合告警归类方法,其特征在于,所述判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合,获得第一判断结果,具体包括:

依次判断多个所述编号日志的规则编号是否存在于所述标号组合中,获得第二判断结果;

若所述第二判断结果为所述编号日志的规则编号存在于所述标号组合中,确定所述编号日志的规则编号位于所述标号组合中的位置;

若所述编号日志的规则编号位于所述标号组合的队首位置,将所述编号日志记录到空队列,生成已匹配队列;

若所述编号日志的规则编号位于所述标号组合的队中位置,判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中,获得第三判断结果;

若所述第三判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中,判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值,获得第四判断结果;

若所述第四判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值,将所述编号日志记录到所述已匹配队列中;

若所述编号日志的规则编号位于所述标号组合的队尾位置,判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中,获得第五判断结果;

若所述第五判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中,判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值,获得第六判断结果;

若所述第六判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值,将所述编号日志记录到所述已匹配队列中,生成完整匹配队列;

根据所述完整匹配队列生成所述第一判断结果,所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合。

3.根据权利要求2所述的日志多组合告警归类方法,其特征在于,所述将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警,具体包括:

将所述完整匹配队列中的多个所述编号日志归类到所述标号组合对应的目标告警事件分类中;

根据所述完整匹配队列中的多个所述编号日志确定告警等级；

根据所述告警等级进行日志事件告警。

4.根据权利要求3所述的日志多组合告警归类方法，其特征在于，所述根据所述完整匹配队列中的多个所述编号日志确定告警等级，具体包括：

获取所述完整匹配队列中的多个所述编号日志的初始风险值及资产价值；

计算所述完整匹配队列中的多个所述编号日志的平均时间间隔；

根据所述初始风险值、所述资产价值及所述平均时间间隔确定风险值；

根据所述风险值确定所述完整匹配队列中的多个所述编号日志的告警等级。

5.一种日志多组合告警归类系统，其特征在于，所述系统包括：

日志获取模块，用于获取多个目标原始日志；多个所述目标原始日志来源于防火墙、网络设备、主机系统、数据库或中间件；

日志过滤模块，用于采用预设的正则表达式过滤多个所述目标原始日志，获得多个过滤后日志；

规则编号添加模块，用于根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储；

告警规则标号获取模块，用于获取目标告警规则中系统动作的标号组合；

匹配模块，用于判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果；

归类告警模块，用于若所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合，将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。

6.根据权利要求5所述的日志多组合告警归类系统，其特征在于，所述匹配模块具体包括：

第二判断子模块，用于依次判断多个所述编号日志的规则编号是否存在于所述标号组合中，获得第二判断结果；

位置确定子模块，用于若所述第二判断结果为所述编号日志的规则编号存在于所述标号组合中，确定所述编号日志的规则编号位于所述标号组合中的位置；

队首日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队首位置，将所述编号日志记录到空队列，生成已匹配队列；

队中日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队中位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第三判断结果；

队中时间判断子模块，用于若所述第三判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第四判断结果；

队中日志记录子模块，用于若所述第四判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中；

队尾日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队尾位

置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第五判断结果；

队尾时间判断子模块，用于若所述第五判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第六判断结果；

队尾日志记录子模块，用于若所述第六判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中，生成完整匹配队列；

第一判断结果生成子模块，用于根据所述完整匹配队列生成所述第一判断结果，所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合。

7.根据权利要求6所述的日志多组合告警归类系统，其特征在于，所述归类告警模块具体包括：

日志归类子模块，用于将所述完整匹配队列中的多个所述编号日志归类到所述标号组合对应的目标告警事件分类中；

告警等级确定子模块，用于根据所述完整匹配队列中的多个所述编号日志确定告警等级；

日志告警子模块，用于根据所述告警等级进行日志事件告警。

8.根据权利要求7所述的日志多组合告警归类系统，其特征在于，所述告警等级确定子模块具体包括：

初始风险值及资产价值获取单元，用于获取所述完整匹配队列中的多个所述编号日志的初始风险值及资产价值；

平均时间间隔计算单元，用于计算所述完整匹配队列中的多个所述编号日志的平均时间间隔；

风险值确定单元，用于根据所述初始风险值、所述资产价值及所述平均时间间隔确定风险值；

告警等级确定单元，用于根据所述风险值确定所述完整匹配队列中的多个所述编号日志的告警等级。

一种日志多组合告警归类方法及系统

技术领域

[0001] 本发明涉及归类分析技术领域，特别是涉及一种日志多组合告警归类方法及系统。

背景技术

[0002] 随着网络技术和网络规模的不断发展，网络系统中的各种网络设备、操作系统、安全设备等都会产生大量的日志数据，为了从海量日志中提取关键数据、提高日志分析效率，提高预警的功能，减少告警漏报、误报的情况，需要对原始日志进行多组合告警归类，从而不用产生无用的告警。只有与设置的多组合告警策略相关的日志，才更有分析和存储价值。因此，将原始日志归类到多组合的告警事件中，保留与告警事件相关的原始日志，舍弃其他无用日志，才能让日志分析人员更加高效的处理这些日志，同时节省日志的存储空间，从而可以存储更长时间的日志。

[0003] 目前，对于日志的归类方法主要采用聚类算法，而聚类算法主要采用统计的方式进行处理，得出的归类结果存在一定程度上的误差，准确性并不是很高，容易存在误判、漏判的情况。

发明内容

[0004] 本发明的目的是提供一种日志多组合告警归类方法及系统，以解决现有日志归类方法归类结果准确性低，且容易存在误判、漏判的问题。

[0005] 为实现上述目的，本发明提供了如下方案：

[0006] 一种日志多组合告警归类方法，所述方法包括：

[0007] 获取多个目标原始日志；多个所述目标原始日志来源于防火墙、网络设备、主机系统、数据库或中间件；

[0008] 采用预设的正则表达式过滤多个所述目标原始日志，获得多个过滤后日志；

[0009] 根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储；

[0010] 获取目标告警规则中系统动作的标号组合；

[0011] 判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果；

[0012] 若所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合，将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。

[0013] 可选的，所述判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果，具体包括：

[0014] 依次判断多个所述编号日志的规则编号是否存在于所述标号组合中，获得第二判断结果；

[0015] 若所述第二判断结果为所述编号日志的规则编号存在于所述标号组合中，确定所

述编号日志的规则编号位于所述标号组合中的位置；

[0016] 若所述编号日志的规则编号位于所述标号组合的队首位置，将所述编号日志记录到空队列，生成已匹配队列；

[0017] 若所述编号日志的规则编号位于所述标号组合的队中位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第三判断结果；

[0018] 若所述第三判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第四判断结果；

[0019] 若所述第四判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中；

[0020] 若所述编号日志的规则编号位于所述标号组合的队尾位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第五判断结果；

[0021] 若所述第五判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第六判断结果；

[0022] 若所述第六判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中，生成完整匹配队列；

[0023] 根据所述完整匹配队列生成所述第一判断结果，所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合。

[0024] 可选的，所述将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警，具体包括：

[0025] 将所述完整匹配队列中的多个所述编号日志归类到所述标号组合对应的目标告警事件分类中；

[0026] 根据所述完整匹配队列中的多个所述编号日志确定告警等级；

[0027] 根据所述告警等级进行日志事件告警。

[0028] 可选的，所述根据所述完整匹配队列中的多个所述编号日志确定告警等级，具体包括：

[0029] 获取所述完整匹配队列中的多个所述编号日志的初始风险值及资产价值；

[0030] 计算所述完整匹配队列中的多个所述编号日志的平均时间间隔；

[0031] 根据所述初始风险值、所述资产价值及所述平均时间间隔确定风险值；

[0032] 根据所述风险值确定所述完整匹配队列中的多个所述编号日志的告警等级。

[0033] 一种日志多组合告警归类系统，所述系统包括：

[0034] 日志获取模块，用于获取多个目标原始日志；多个所述目标原始日志来源于防火墙、网络设备、主机系统、数据库或中间件；

[0035] 日志过滤模块，用于采用预设的正则表达式过滤多个所述目标原始日志，获得多个过滤后日志；

[0036] 规则编号添加模块，用于根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储；

[0037] 告警规则标号获取模块，用于获取目标告警规则中系统动作的标号组合；

[0038] 匹配模块，用于判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果；

[0039] 归类告警模块，用于若所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合，将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。

[0040] 可选的，所述匹配模块具体包括：

[0041] 第二判断子模块，用于依次判断多个所述编号日志的规则编号是否存在于所述标号组合中，获得第二判断结果；

[0042] 位置确定子模块，用于若所述第二判断结果为所述编号日志的规则编号存在于所述标号组合中，确定所述编号日志的规则编号位于所述标号组合中的位置；

[0043] 队首日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队首位置，将所述编号日志记录到空队列，生成已匹配队列；

[0044] 队中日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队中位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第三判断结果；

[0045] 队中时间判断子模块，用于若所述第三判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第四判断结果；

[0046] 队中日志记录子模块，用于若所述第四判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中；

[0047] 队尾日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队尾位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第五判断结果；

[0048] 队尾时间判断子模块，用于若所述第五判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第六判断结果；

[0049] 队尾日志记录子模块，用于若所述第六判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中，生成完整匹配队列；

[0050] 第一判断结果生成子模块，用于根据所述完整匹配队列生成所述第一判断结果，所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合。

[0051] 可选的，所述归类告警模块具体包括：

[0052] 日志归类子模块，用于将所述完整匹配队列中的多个所述编号日志归类到所述标号组合对应的目标告警事件分类中；

[0053] 告警等级确定子模块，用于根据所述完整匹配队列中的多个所述编号日志确定告

警等级；

[0054] 日志告警子模块，用于根据所述告警等级进行日志事件告警。

[0055] 可选的，所述告警等级确定子模块具体包括：

[0056] 初始风险值及资产价值获取单元，用于获取所述完整匹配队列中的多个所述编号日志的初始风险值及资产价值；

[0057] 平均时间间隔计算单元，用于计算所述完整匹配队列中的多个所述编号日志的平均时间间隔；

[0058] 风险值确定单元，用于根据所述初始风险值、所述资产价值及所述平均时间间隔确定风险值；

[0059] 告警等级确定单元，用于根据所述风险值确定所述完整匹配队列中的多个所述编号日志的告警等级。

[0060] 根据本发明提供的具体实施例，本发明公开了以下技术效果：

[0061] 本发明提供一种日志多组合告警归类方法及系统，所述方法首先获取多个目标原始日志；采用预设的正则表达式过滤多个所述目标原始日志，获得多个过滤后日志；并根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储；然后获取目标告警规则中系统动作的标号组合；判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，若是，将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。本发明提供的方法通过判断目标日志的规则编号是否能够完整匹配系统动作的标号组合，而不仅仅匹配一个动作，从而避免了现有的日志归类过程中，归类结果准确性不高，且存在误判、漏判的问题，提高了日志归类、告警的准确性。

附图说明

[0062] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例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发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0063] 图1为本发明提供的日志多组合告警归类方法的方法流程图；

[0064] 图2为本发明提供的匹配组合告警日志的方法流程图；

[0065] 图3为本发明提供的采用决策树算法挖掘告警规则的方法流程图；

[0066] 图4为本发明提供的确定告警等级的方法流程图；

[0067] 图5为本发明提供的日志多组合告警归类系统的系统结构图。

具体实施方式

[0068] 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，所描述的实施例仅仅是本发明一部分实施例，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前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0069] 本发明的目的是提供一种日志多组合告警归类方法及系统，以解决现有日志归类方法归类结果准确性低，且容易存在误判、漏判的问题。

[0070] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0071] 图1为本发明提供的日志多组合告警归类方法的方法流程图。参见图1，本发明提供的日志多组合告警归类方法具体包括：

[0072] 步骤101：获取多个目标原始日志。

[0073] 获取目标原始日志有主动采集方式和被动接受方式，主动采集方式针对windows主机设备采用远程桌面方式连接，获取系统日志、数据库、中间件日志，针对linux主机设备采用ssh方式连接，获取系统、数据库、中间件日志。

[0074] 针对不可提供主动采集的设备，可采用被动采集方式，设备配置syslog或安装可采集日志的客户端，将日志数据统一发往日志接收服务器，从而获取所述目标原始日志。

[0075] 多个所述目标原始日志来源于防火墙、网络设备、主机系统、数据库或中间件。

[0076] 步骤102：采用预设的正则表达式过滤多个所述目标原始日志，获得多个过滤后日志。

[0077] 步骤103：根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储。

[0078] 日志过滤模块将所述步骤101中获取到的目标原始日志，通过预设的正则表达式进行过滤，提取目标原始日志中的敏感日志，并将提取的过滤后日志转换成统一格式，并添加正则表达式规则编号，生成编号日志存储在数据库中，用于后续进行多组合告警处理。

[0079] 提取的编号日志的格式为：

[0080] [时间][ip地址][日志类型][规则编号][原始日志]

[0081] 例如格式为“1522639541c0a80102010101d1c2m6z0ty00192.168.138.1000SType12AXTXId＝6013AXTXSrc＝Eventlog AXTXMsg＝系统启动时间为20秒”的编号日志，其中“1522639541”为时间，“c0a80102”为对应的ip地址，“0”为日志类型，这里0代表是系统日志，“101”为对应的规则编号，其余为日志原始数据。所述日志类型包括系统日志(用0表示)，应用日志(用1表示)，其他日志(用2表示)。

[0082] 所述步骤102主要是过滤目标原始日志，匹配所设过滤规则；所述步骤103主要是为过滤后日志添加正则表达式的规则编号，因为后续对目标原始日志进行匹配都是基于正则表达式的规则编号。所述规则编号是系统动作库中对应动作在数据库中的ID编号。系统动作可能是登录、刷新等操作，每个这种操作都是一个动作。

[0083] 步骤104：获取目标告警规则中系统动作的标号组合。

[0084] 根据预设的目标告警规则，获取目标告警规则中系统动作的标号组合。本发明中每个目标告警规则都是由多个系统动作组成的，所述目标告警规则中系统动作的标号组合就是这个目标告警规则中所有的系统动作在数据库中对应的ID编号的组合。

[0085] 设置所述目标告警规则可以有三种形式：

[0086] 一种为内置规则，内置规则是根据已有的漏洞去添加，这样当发布一种新的漏洞时，就可以及时进行添加从而可以避免不必要的损失。

[0087] 第二种为自定义规则，自定义规则根据已有的大量日志数据，利用数据挖掘算法去隐藏安全事件，也可以设置比较确定的规则，如“root用户登录”，这样只有root用户登录时才会产生告警，而不是所有用户登录都会产生告警，从而实现可配置的低误报，同时也可

以起到预警的作用。

[0088] 另外还有第三种告警规则，第三种告警规则是统计已有的日志，依据统计结果计算出各个系统动作发生时产生告警的概率，再利用条件概率计算出当任一系统动作发生时，其他系统动作发生时会产生告警的概率。根据统计的结果，利用机器学习从已有的数据中挖掘出潜在的一系列危险的系统动作形成相应的告警规则，自动添加到告警规则中从而可以起到提前预警的作用。本发明优选采用第三种告警规则。

[0089] 步骤105：判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果。

[0090] 所述步骤105用于检索由所述步骤103产生的编号日志，同时基于所述步骤104中获取的目标告警规则中系统动作的标号组合，判断所检索的编号日志的规则编号是否能够完整匹配所述标号组合中的每个ID编号。

[0091] 所述步骤105连续从所述步骤103中获取编号日志，即每次用最近事件产生的日志数据去匹配步骤104中的标号组合。若匹配成功，即若能够完整匹配，则将构成目标告警规则中系统动作的标号组合的日志归类到对应的分类中。

[0092] 图2为本发明提供的匹配组合告警日志的方法流程图。参见图2，所述判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果，具体包括：

[0093] 步骤S41：获取过滤后的目标日志，即所述步骤103所记录的编号日志，设置周期获取最新产生的编号日志，如最近一个周期内无新的编号日志产生，等待下一周期执行。

[0094] 步骤S42：获取组合告警动作编号，即所述步骤104中的目标告警规则中系统动作的标号组合。可设置告警开关，用于开启和关闭组合告警状态。当获取所述标号组合时，只获取数据库中预先存储的处于开启状态的目标告警规则中系统动作的标号组合，对于处于关闭状态的告警标号组合，获取时将忽略。如获取标号组合顺序为a_1-a_n:[a_1,a_2,a_3,.....,a_n]，其中的a_1-a_n分别代表所述标号组合对应的规则编号。

[0095] 步骤S43：判断目标日志规则编号是否在组合告警编号内，即判断所述编号日志的规则编号是否存在于所述标号组合中，解析步骤S41获取的编号日志，得到所述编号日志的规则编号，判断所述规则编号是否在步骤S42获取的标号组合中，若否，忽略该条编号日志，返回所述步骤S41，获取下一条编号日志。

[0096] 步骤S44：计算目标日志规则编号位置，即确定所述编号日志的规则编号位于所述标号组合中的位置；对于步骤S43，若是，匹配该条编号日志的规则编号位于所述标号组合中的位置，所述标号组合为目标告警规则中系统动作对应的ID编号的列表。所述规则编号位于所述标号组合中的位置存在三种情况，位于队列首位，位于队列中位，位于队列末位，三种位置处理步骤分别对应步骤S45、S46、S47。

[0097] 步骤S45：记录告警规则位于组合首位的日志，即对于所述规则编号位于所述标号组合的队首位置的编号日志，新建一个空队列，将位于队列首位的编号日志存储至新建队列中，生成已匹配队列。

[0098] 例如该条日志规则编号为a_1，新建队列L，已匹配队列为：

[0099] L＝[a_1,]

[0100] 步骤S46：记录告警规则位于组合中位的日志，及对于所述规则编号位于所述标号组合的队中位置的编号日志，查找现有的已匹配队列中是否存在前一规则编号的编号日

志，若存在，计算前一规则编号对应的编号日志与当前判断的编号日志的时间间隔是否超过设置的阈值，若是，舍弃当前判断的所述编号日志；若否，将当前判断的所述编号日志存储到已匹配队列中；若所述已匹配队列中已存在相同规则编号的日志，则将已匹配队列中已有的日志更新为当前判断的该条编号日志。

[0101] 例如当前判断的该条编号日志的规则编号为a_2，判断已匹配队列L中是否存在前一条规则编号对应的编号日志a_1，若已匹配队列L中存在a_1，则计算a_1到a_2的时间间隔t，若t＜阈值，则将a_2存储到已匹配队列L中，此时已匹配队列更新为：

[0102] $L=[a_1,a_2]$

[0103] 步骤S47：记录告警规则位于组合末位的日志，同所述步骤S46，将处于末位的编号日志存储至队列。具体为：若所述编号日志的规则编号位于所述标号组合的队尾位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，若是，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，若是，将所述编号日志记录到所述已匹配队列中，生成完整匹配队列。

[0104] 例如对于告警规则位于末位的日志a_n，生成完整匹配队列L：

[0105] $L=[a_1,a_2,a_3,\ldots\ldots,a_n]$

[0106] 步骤S48：告警。

[0107] 此时所述完整匹配队列L中已完整记录了所述标号组合中ID编号为a_1-a_n的日志，表示多个所述编号日志的规则编号能够完整匹配所述标号组合，此时将根据所述完整匹配队列L中记录的编号日志产生告警。

[0108] 步骤106：若所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合，将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警。具体包括：

[0109] 步骤(1)：将所述完整匹配队列中的多个所述编号日志归类到所述标号组合对应的目标告警事件分类中。

[0110] 步骤(2)：根据所述完整匹配队列中的多个所述编号日志确定告警等级。

[0111] 根据所述步骤105匹配到的结果，确认对应的告警等级，从而以低、中、高三个等级进行告警。主要是根据组合告警动作的初始风险值及资产价值、构成组合告警动作的目标日志的总事件平均间隔来计算出产生的组合告警的风险值，从而确定是否告警以及告警的等级，若告警等级太低，则说明此条告警的价值不是很高，可以不用告警，从而减少误报的可能。

[0112] 从已经产生告警的日志样本数据(即所述完整匹配队列中的多个所述编号日志)中，可以得出产生告警的系统动作编号为$[a_1,a_2,a_3,\ldots\ldots,a_n]$，$[a_1,a_2,a_3,\ldots\ldots,a_n]$即为所述完整匹配队列中的多个所述编号日志对应的规则编号。在这里a_i表示所述完整匹配队列中第i个编号日志对应的规则编号，a_i具体代表的是一个系统动作，在本发明中系统动作就是在进行操作时的一个动作，例如“点击刷新按钮”就是一个动作。

[0113] 根据已获取的样本数据分别计算系统动作a_i发生时产生告警的概率$p(a_i)$，$i\in[1,n]$，如公式(1)所示：

[0114] $$p(a_i)=\frac{N_i}{S} \quad (1)$$

[0115] 其中N_i是系统动作a_i产生告警的次数，S是产生告警的总次数，n为系统动作a_i的数量。

[0116] 本发明不仅仅要分析一个系统动作产生的概率，还要通过已有的告警规则挖掘出潜在的有危险的告警规则，因此还要分析系统动作之间的关系。通过条件概率公式来计算当系统动作a_i产生时，其余系统动作a_j产生的概率，其中$j\in[1,n]$，$j\neq i$，这里用$p(a_j|a_i)$表示，如公式(2)所示：

[0117] $$p(a_j|a_i)=p(a_ja_i)/p(a_i) \qquad (2)$$

[0118] 其中$p(a_ja_i)$表示a_j和a_i按照a_j在前的顺序出现时，产生告警的概率，和所述$p(a_i)$的计算方法类似，根据已有的告警日志进行计算。

[0119] 采用所述公式(1)、(2)进行概率计算时，由于初始样本数据的局限性，可能导致概率计算不准确，可在后续获取大量告警日志数据后不断的优化使样本概率逐渐趋向稳定。通过计算出的系统动作之间的关系，就可以利用数据挖掘中的算法将潜在有危险的一系列动作挖掘出来形成相应的告警规则。

[0120] 现有的数据挖掘算法有很多，而本发明要根据多个特征发生的概率得出可能存在风险的告警规则，很显然符合决策树的应用条件。因此，本发明采用决策树来得到想要的结果。决策树算法主要包括三个部分：特征的选择，树的生成、树的剪枝。

[0121] 特征选择：目的是选取能够对训练集分类的特征。特征选择的关键是准则，一般根据信息增益、信息增益比、Gini指数进行选择，本发明采用信息增益比。

[0122] 决策树的生成：通常是利用信息增益最大、信息增益比最大、Gini指数最小作为特征选择的准则。从根结点开始，递归的生成决策树。

[0123] 决策树的剪枝：决策树的剪枝是为了防止树的过拟合，增强其泛化能力。包括预剪枝和后剪枝。

[0124] 本发明将需要分析的多个编号日志作为训练集D，而每个系统动作a_i就是一个特征值，在本发明中主要是挖掘潜在有危险的一系列的系统动作形成相应的告警规则。因此，选取哪个特征值成为根结点，就需要计算每个特征值对应的信息增益比。这里就需要引入一些概念。

[0125] 首先引入熵的概念，熵H定义为信息的增益值，用来衡量随机变量的不确定性，通过公式(3)得到：

[0126] $$H=-\sum_{i=1}^{n}p(a_i)\log_2 p(a_i) \qquad (3)$$

[0127] 当熵中的概率由日志数据估计得到时，所对应的熵称为经验熵。这里训练数据集D的经验熵为H(D)，|D|表示样本容量，即样本个数。设有K个类C_k，其中$k=1,2,3,\cdots,K$，$|C_k|$为属于类C_k的样本个数，则利用公式(4)来计算经验熵：

[0128] $$H(D)=-\sum\frac{|C_k|}{D}\log_2\frac{|C_k|}{|D|} \qquad (4)$$

[0129] 在计算信息增益比之前，还要知道条件熵的概念。条件熵H(Y|X)表示在已知随机变量X的条件下随机变量Y的不确定性，随机变量X给定的条件下随机变量Y的条件熵，由公式(5)表示：

[0130] $$H(Y|X)=\sum_{i=1}^{m}p_iH(Y|X=x_i) \qquad (5)$$

[0131] 其中p_i是当随机变量为x_i时的概率，$p_i=P(X=x_i)$，x_i就是随机变量，m是所有随机变量的数目；此时如果有0的概率，令0log0＝0。这时，再计算信息增益，信息增益是相对于特征而言的。

[0132] 特征a_i对训练数据集D的信息增益$g(D,a_i)$，定义为集合D的经验熵H(D)与给定特征a_i条件下训练数据集D的经验条件熵$H(D|a_i)$之差，即如公式(6)所示：

[0133] $$g(D,a_i)=H(D)-H(D|a_i) \qquad (6)$$

[0134] 这里利用信息增益比来选取特征，特征a_i对训练数据集D的的信息增益比$g_R(D,a_i)$，定义为其信息增益$g(D,a_i)$与训练数据集D的经验熵之比，如公式(7)所示：

[0135] $$g_R=\frac{g(D,a_i)}{H(D)} \qquad (7)$$

[0136] 利用公式(7)求得的值，可以进行决策树的构建，本发明采用的是C4.5算法，具体的流程如图2所示。

[0137] 图3为本发明提供的采用决策树算法挖掘告警规则的方法流程图。本发明采用的决策树算法输入为训练数据集D、特征集A和阈值ε；输出为决策树T。所述特征集A是特征a_i的集合。参见图3，本发明提供的采用决策树算法挖掘告警规则的方法包括：

[0138] 步骤S31：判断所有实例是否都属于同一类C_k；如果D中所有实例属于同一类C_k，则置T为单结点树，并将C_k作为该结点的类，返回T；

[0139] 步骤S32：判断特征集是否为空；如特征集$A=\varnothing$，则置T为单结点树，并将D中实例数最大的类C_k作为该结点的类，返回T；

[0140] 步骤S33：计算特征集的信息增益比；若特征集不为空，则计算A中各特征对D的信息增益比，选择信息增益比最大的特征$A_g=a_i$；

[0141] 步骤S34：判断信息增益比最大的特征A_g是否小于阈值；如果A_g的信息增益比小于阈值ε，则置T为单结点树，并将D中实例数最大的类C_k作为该结点的类，返回T；

[0142] 步骤S35：计算每一子结点；如果A_g的信息增益比不小于阈值ε，则对A_g的每一可能值a_j(其中$j\neq i$)，将D分割成若干非空子集D_j，计算每一子结点，以D_j为训练集，以$A-\{a_i,a_j\}$为特征集，递归的调用步骤S31～步骤S34，由结点及其子结点构成树T，返回T；直至$A-\{a_i,a_j\}$为空，递归调用过程结束。

[0143] 决策树算法很容易过拟合，剪枝算法就是用来防止决策树过拟合，提高泛化性能的方法。本发明采用后剪枝算法，后剪枝是指先从训练集生成一棵完整的决策树，然后自底向上对非叶结点进行考察，若将该结点对应的子树替换为叶结点，能带来泛化性能的提升，则将该子树替换为叶结点。

[0144] 最后生成树的每个根结点到叶子结点是一条挖掘出的有潜在风险的告警规则，每个结点都是一个系统动作。然后，将挖掘出的有潜在风险的告警规则作为预测出的告警规则添加到系统的告警规则库中，从而更新所述目标告警规则。

[0145] 所述步骤(2)根据所述完整匹配队列中的多个所述编号日志确定告警等级。

[0146] 本发明实施例中，当日志中存在目标告警动作标号的组合，则要根据组合告警动作的初始风险值、资产的价值、构成组合告警动作的目标日志的总事件平均间隔来计算出

产生的组合告警的风险值，从而确定是否告警以及告警的等级，若告警等级太低，则说明此条告警的价值不是很高，可以不用告警，从而减少误报的可能。

[0147] 为了能够确定告警的等级，本发明实施例中，将计算出组合告警的风险值进行等级的划分，并根据相应的等级在告警的时候，展示给用户的是高中低3个等级。

[0148] 图4为本发明提供的确定告警等级的方法流程图，参见图4，本发明实施例中所述步骤(2)确定告警等级包括步骤：

[0149] S51、确认组合告警动作初始风险值。

[0150] 根据已有的漏洞库和各个漏洞可能对主机系统、设备造成的危险程度来确定组合告警动作初始风险值，当漏洞对系统造成了不可修复即毁灭性的危害时，这个漏洞对应的组合告警动作的初始风险值将是最高的。

[0151] 本发明用d_i来表示第i个所述完整匹配队列的初始风险值，如果其中某个组合告警动作已经有通过打补丁或者其余的方式进行避免，则系统也会降低其对应的风险值，风险值的后续变化将会利用机器学习，从而挖掘动作之间的联系。并且d_i满足的条件如公式(8)所示：

[0152] $0 \leqslant d_i \leqslant 1$ (8)

[0153] S52、确定资产的价值；

[0154] 本发明实施例中，主要是对日志分析的系统，所以本发明根据资产在单位时间内T_0发送的日志量，来判断资产的重要性。将根据所有资产在单位时间T_0内所发的日志量进行排序，并以1～10来确定每个资产的价值，用V_n来表示第n个资产的资产价值，如公式(9)所示：

[0155] $1 \leqslant V_n \leqslant 10$ (9)

[0156] S53、计算构成组合告警策略的目标日志总事件的平均间隔；即，计算所述完整匹配队列中的多个所述编号日志的平均时间间隔。

[0157] 本发明实施例中，因为是组合告警动作，所以是由多个动作依据顺序全部完成之后，才会告警。但是，如果两个动作之间的时间间隔太大，本发明会认为这两个动作不是按顺序完成，因此不会归为同一组合告警动作，即使发生也不会告警。

[0158] 本发明用t_{ij}表示第i个完整匹配队列中第j个动作和第j+1个动作的时间间隔，其中$j \geqslant 1$。因为每个完整匹配队列中的动作数是不一样的，因此这里计算每个完整匹配队列中所有动作时间间隔的平均值，即假设第i个完整匹配队列中有j个动作，则这条完整匹配队列的平均时间间隔如公式(10)所示：

[0159] $$\bar{t}_i = \frac{t_{i1} + t_{i2} + t_{i3} + \cdots + t_{i(j-1)}}{j-1} \qquad (10)$$

[0160] 本发明设置所述平均时间间隔的间隔阈值为1个小时，如果平均时间间隔大于1个小时，则不会告警。这里用0表示不告警，如公式(11)所示：

[0161] $$t_i = \begin{cases} 0, & \bar{t}_i > 1 \\ \bar{t}_i, & 0 < \bar{t}_i \leq 1 \end{cases} \qquad (11)$$

[0162] S54、计算产生的组合告警的风险值；即，根据所述初始风险值、所述资产价值及所述平均时间间隔确定所述完整匹配队列的风险值。

[0163] 本发明实施例中，将根据所述完整匹配队列的初始风险值及资产价值和平均时间间隔来计算产生的组合告警的风险值，如公式(12)所示，第n个资产产生第i条完整匹配队列的风险值为：

[0164] $R_{ni}=d_i\times V_n\times t_i$ (12)

[0165] 这里，当产生第i条完整匹配队列时，会根据日志内IP地址字段来判断是哪台资产，从而找到对应的资产价值。

[0166] S55、确认告警等级；即根据所述风险值确定所述完整匹配队列中的多个所述编号日志的告警等级。

[0167] 依据公式(8)～公式(12)计算出第n个资产发生第i条组合告警动作时的风险值R_{ni}，这里就要依据风险值来确定对应的告警等级。由公式(12)可以得出R_{ni}的范围，如公式(13)所示：

[0168] $0\leqslant R_{ni}\leqslant 10$ (13)

[0169] 本发明设置3个告警等级，分别为低、中、高。因此采用表一中的范围划分来确认对应的告警等级。

[0170] 表一

[0171]

完整匹配队列的风险值	告警等级
0-3.33	低

[0172]

3.34-6.66	中
6.67-10	高

[0173] 根据得出的告警等级进行对应的告警，这里告警的方式也可以根据告警等级来设置。如果告警等级处于低或者中的等级，则可以通过界面的形式在首页展示，如果告警等级为高时，可以通过发送邮件或者短信的形式第一时间通知管理人员，及时解决问题，以免造成不可挽回的损失。

[0174] 本发明实施例中，与所述的日志多组合告警归类方法相对应的，本发明还提供了一种日志多组合告警归类系统，图5为本发明提供的日志多组合告警归类系统的系统结构图，参见图5，所述系统包括：

[0175] 日志获取模块501，用于获取多个目标原始日志；多个所述目标原始日志来源于防火墙、网络设备、主机系统、数据库或中间件；

[0176] 日志过滤模块502，用于采用预设的正则表达式过滤多个所述目标原始日志，获得多个过滤后日志；

[0177] 规则编号添加模块503，用于根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储；

[0178] 告警规则标号获取模块504，用于获取目标告警规则中系统动作的标号组合；

[0179] 匹配模块505，用于判断多个所述编号日志的规则编号是否能够完整匹配所述标号组合，获得第一判断结果；

[0180] 归类告警模块506，用于若所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合，将多个所述编号日志归类到所述标号组合对应的目标告警事件

分类中并产生告警。

[0181] 其中，所述匹配模块505具体包括：

[0182] 第二判断子模块，用于依次判断多个所述编号日志的规则编号是否存在于所述标号组合中，获得第二判断结果；

[0183] 位置确定子模块，用于若所述第二判断结果为所述编号日志的规则编号存在于所述标号组合中，确定所述编号日志的规则编号位于所述标号组合中的位置；

[0184] 队首日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队首位置，将所述编号日志记录到空队列，生成已匹配队列；

[0185] 队中日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队中位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第三判断结果；

[0186] 队中时间判断子模块，用于若所述第三判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第四判断结果；

[0187] 队中日志记录子模块，用于若所述第四判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中；

[0188] 队尾日志处理子模块，用于若所述编号日志的规则编号位于所述标号组合的队尾位置，判断所述标号组合中所述编号日志的规则编号的前一规则编号是否已在所述已匹配队列中，获得第五判断结果；

[0189] 队尾时间判断子模块，用于若所述第五判断结果为所述编号日志的规则编号的前一规则编号已在所述已匹配队列中，判断所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔是否小于预设的阈值，获得第六判断结果；

[0190] 队尾日志记录子模块，用于若所述第六判断结果为所述编号日志与所述前一规则编号对应的前一编号日志的时间间隔小于预设的阈值，将所述编号日志记录到所述已匹配队列中，生成完整匹配队列；

[0191] 第一判断结果生成子模块，用于根据所述完整匹配队列生成所述第一判断结果，所述第一判断结果为多个所述编号日志的规则编号能够完整匹配所述标号组合。

[0192] 所述归类告警模块506具体包括：

[0193] 日志归类子模块，用于将所述完整匹配队列中的多个所述编号日志归类到所述标号组合对应的目标告警事件分类中；

[0194] 告警等级确定子模块，用于根据所述完整匹配队列中的多个所述编号日志确定告警等级；

[0195] 日志告警子模块，用于根据所述告警等级进行日志事件告警。

[0196] 所述告警等级确定子模块具体包括：

[0197] 初始风险值及资产价值获取单元，用于获取所述完整匹配队列中的多个所述编号日志的初始风险值及资产价值；

[0198] 平均时间间隔计算单元，用于计算所述完整匹配队列中的多个所述编号日志的平均时间间隔；

[0199] 风险值确定单元，用于根据所述初始风险值、所述资产价值及所述平均时间间隔确定风险值；

[0200] 告警等级确定单元，用于根据所述风险值确定所述完整匹配队列中的多个所述编号日志的告警等级。

[0201] 在实际应用中，本发明提供的系统还可以包括：

[0202] 日志采集模块，部署在需要采集日志的主机、主机系统、网络设备、数据库、中间件等上，用于采集设备的系统日志和应用日志。

[0203] 数据库模块，用于保存所述日志采集模块、日志获取模块501、日志过滤模块502、规则编号添加模块503、告警规则标号获取模块504、匹配模块505以及归类告警模块506所需要的系统动作、资产数据、策略信息和漏洞信息。

[0204] 系统动作库模块，用于添加用于组合关联告警的动作，将动作统一进行添加和保存，并在动作中添加用于日志过滤的正则表达式，从而将原始日志数据进行过滤。

[0205] 告警规则过滤器模块，用于暂时保存认为没有威胁的告警策略，并用于查看相应告警策略产生的日志。

[0206] 本发明提供的方法及系统通过匹配连续出现的系统动作的标号组合，可以将目标日志归类到对应的告警事件中，避免了现有的日志归类过程中，归类结果单一、繁多，人为识别组合出现的漏判问题。

[0207] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

[0208] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处。综上所述，本说明书内容不应理解为对本发明的限制。

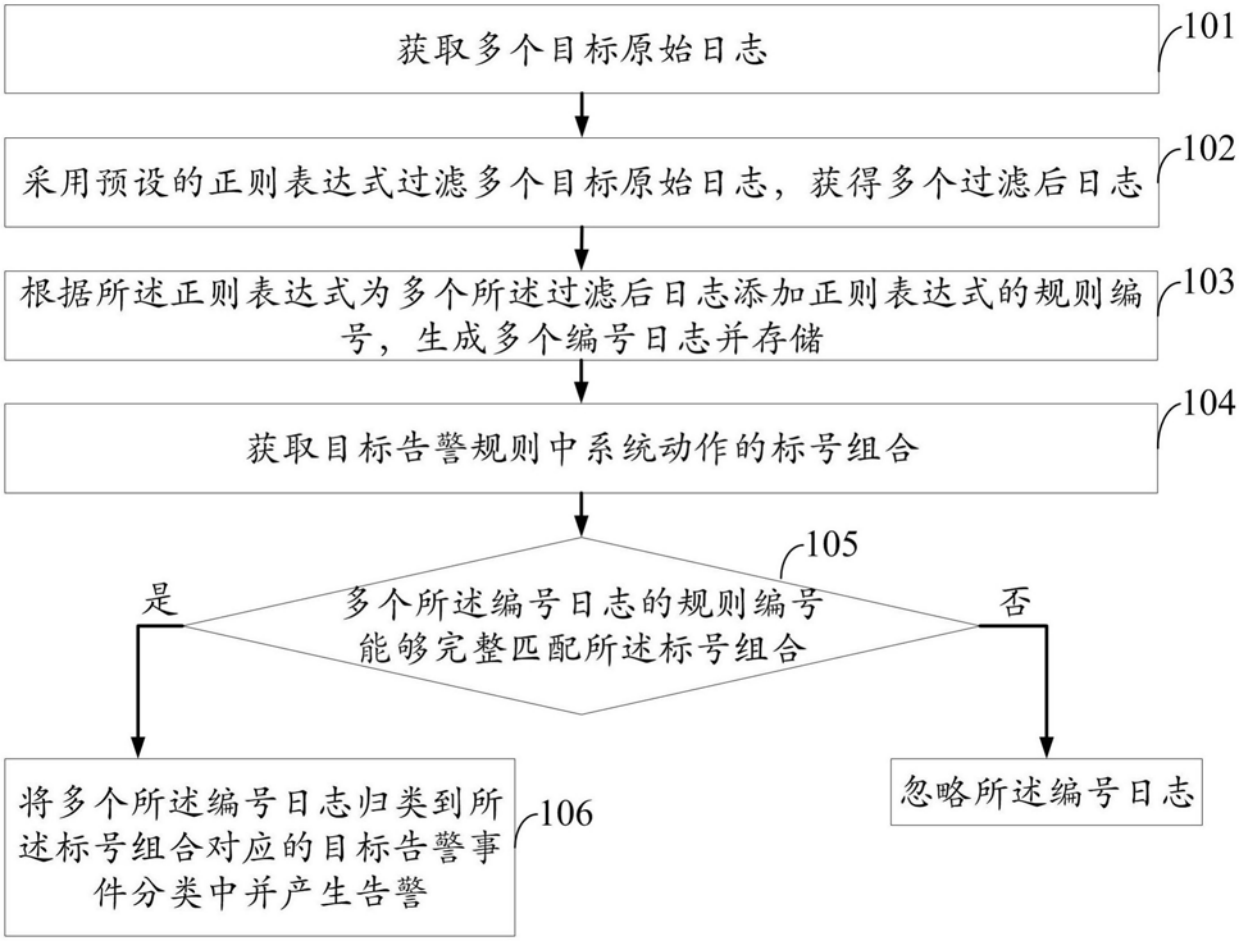

获取多个目标原始日志
101
采用预设的正则表达式过滤多个目标原始日志，获得多个过滤后日志
102
根据所述正则表达式为多个所述过滤后日志添加正则表达式的规则编号，生成多个编号日志并存储
103
获取目标告警规则中系统动作的标号组合
104
多个所述编号日志的规则编号能够完整匹配所述标号组合
105
是
否
将多个所述编号日志归类到所述标号组合对应的目标告警事件分类中并产生告警
106
忽略所述编号日志

图1

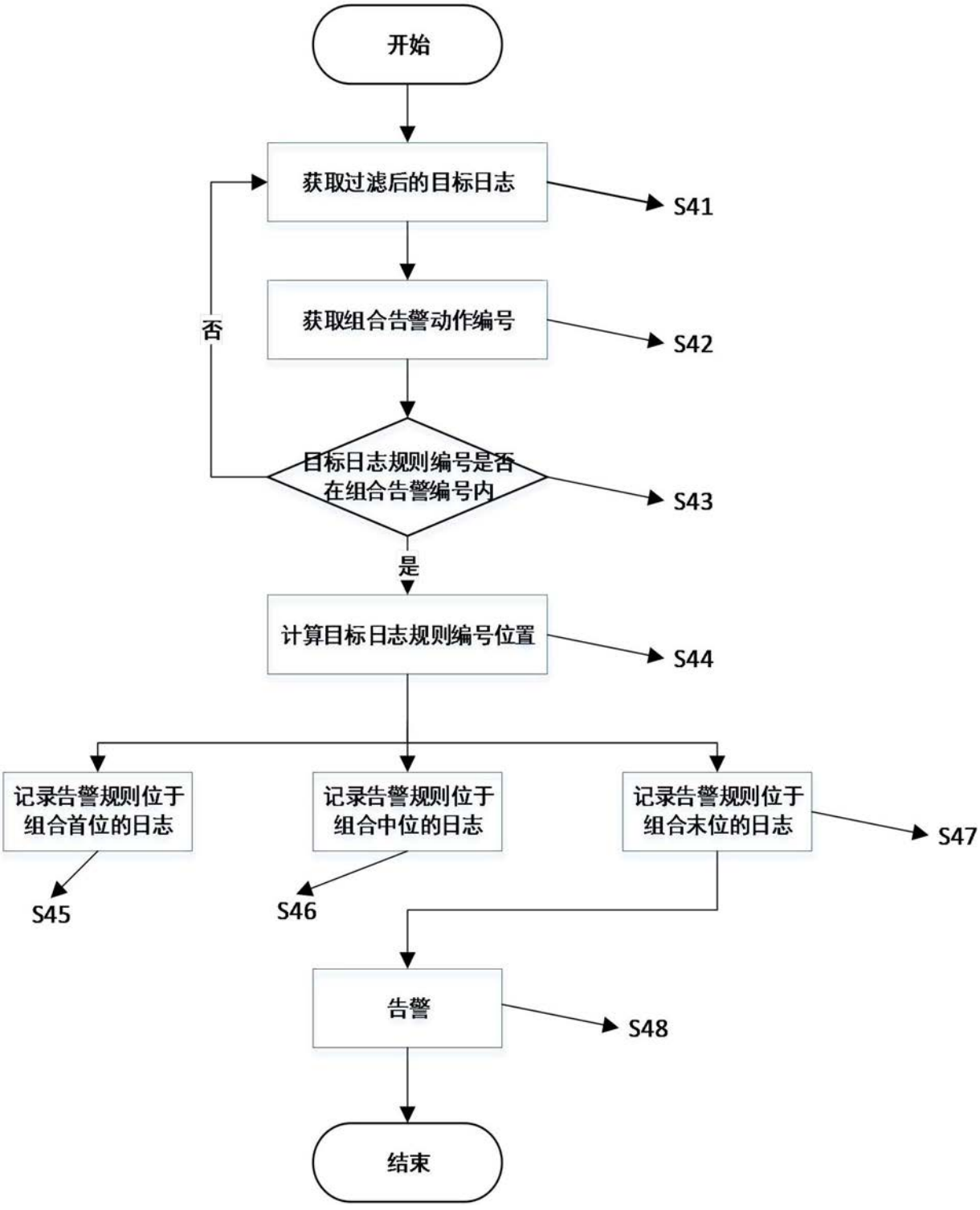

开始
获取过滤后的目标日志
S41
获取组合告警动作编号
S42
否
目标日志规则编号是否在组合告警编号内
S43
是
计算目标日志规则编号位置
S44
记录告警规则位于组合首位的日志
记录告警规则位于组合中位的日志
记录告警规则位于组合末位的日志
S47
S45
S46
告警
S48
结束

图2

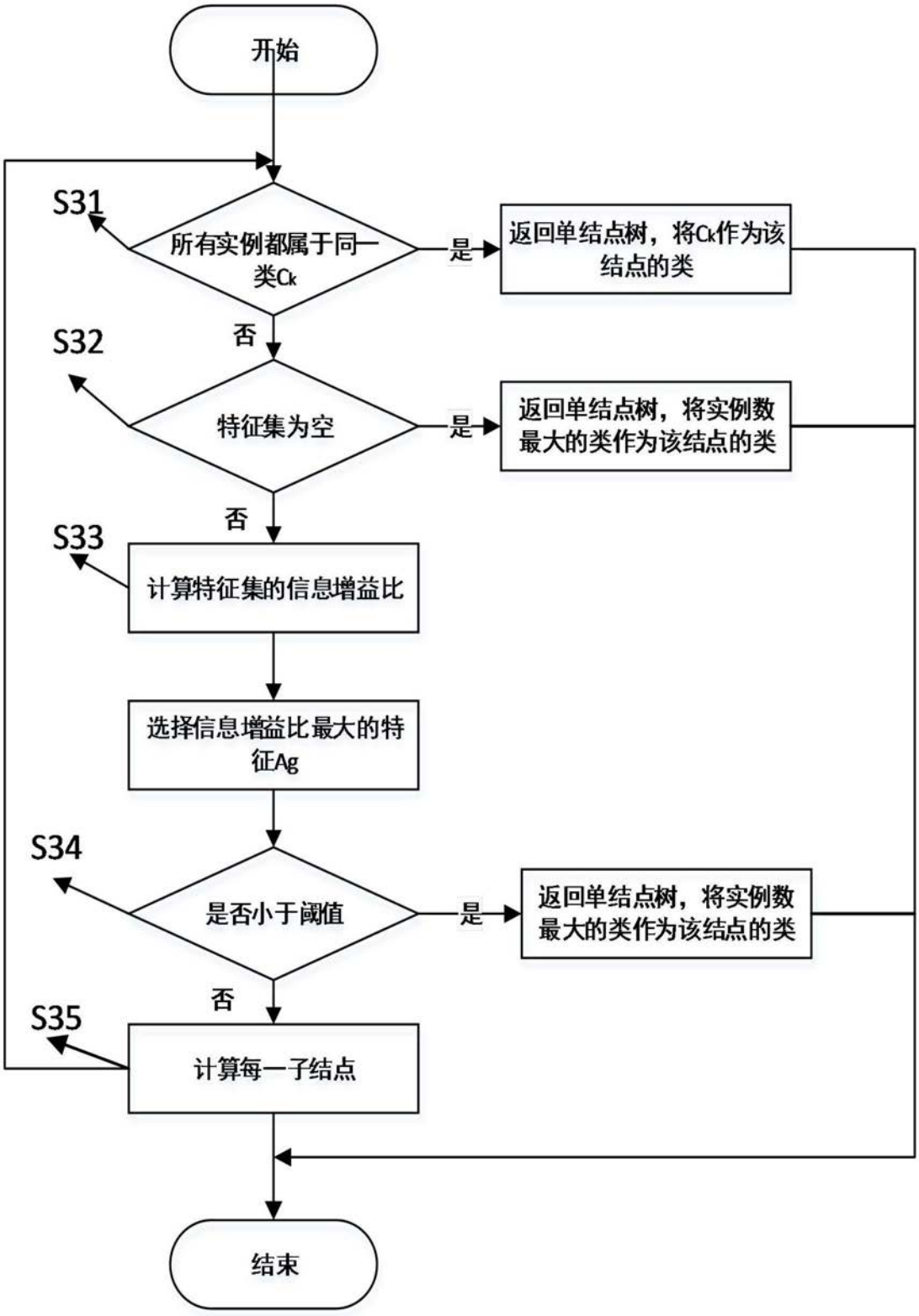

开始
S31
所有实例都属于同一类Ck
是
返回单结点树，将Ck作为该结点的类
否
S32
特征集为空
是
返回单结点树，将实例数最大的类作为该结点的类
否
S33
计算特征集的信息增益比
选择信息增益比最大的特征Ag
S34
是否小于阈值
是
返回单结点树，将实例数最大的类作为该结点的类
否
S35
计算每一子结点
结束

图3

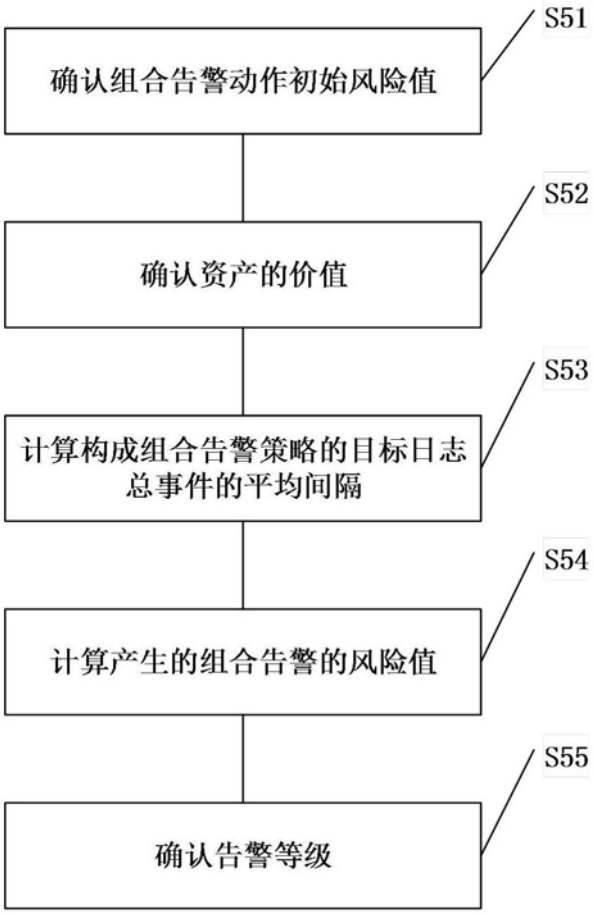

S51
确认组合告警动作初始风险值
S52
确认资产的价值
S53
计算构成组合告警策略的目标日志总事件的平均间隔
S54
计算产生的组合告警的风险值
S55
确认告警等级

图4

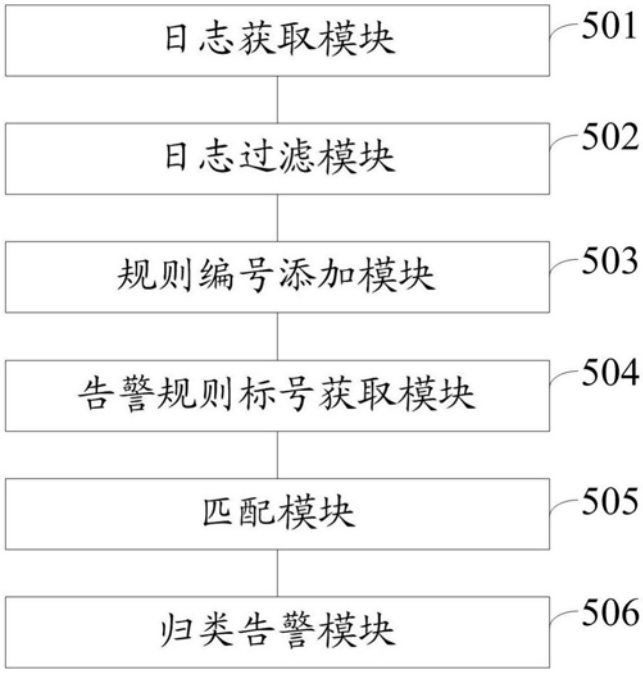

日志获取模块
501
日志过滤模块
502
规则编号添加模块
503
告警规则标号获取模块
504
匹配模块
505
归类告警模块
506

图5