#### **标题：一种基于日志关联挖掘的威胁检测方法**

##### 

（1）数据收集

本发明以日志信息作为输入，本步骤旨在从系统日志中提取所需的结构化数据，设日志条目列表可以记为

单个日志条目通常由时间戳、日志ID和详细日志信息组成，总体可以表示为

其中详细日志信息包含多个单词，根据日志种类的不同可以划分为进程、文件、网络地址等多种对象信息，又根据语义不同可以划分为源和目标两类，因而存在一种映射，使得对于每种日志类型，可以将进一步表示为

其中表示为日志的源对象（集合），表示为日志的目标对象（集合），表示日志类型，为保留的日志上下文信息。映射后的日志信息可以表示为

为了便于后续分析和规则挖掘，对于源对象和目标对象进行一定层次的抽象，如from可以分为外部来源、系统用户、管理员用户、系统进程、脚本等。

例如，对于/var/log/secure的一条日志记录

*Dec 27 14:04:53 139 sshd[30968]: Invalid user user from 193.201.224.12 port 26133*

解析如下，timestamp为*Dec 27 14:04:53，*id由系统自动生成，from为外部来源（193.201.224.12:26133），to为sshd，type根据日志记录的名称即secure，ext\_info为invalid user。

根据MITRE提出的ATT&CK模型，该模型根据真实的观察数据来描述和分类对抗行为，目标是创建网络攻击中使用的已知对抗战术和技术的详尽列表。该模型给出了由攻击者在攻击企业时会利用的244种企业技术组成的精选知识库构成的“对抗战术、技术和常识”框架。

根据该模型，我们可以通过NLP或者手动的方式构建由日志到模型的映射关系，添加代表映射的模型，如Untrusted Read, Untrusted Exec等映射后的日志信息可以表示为

（2）日志依赖关系构建

将前述的每条结构化的日志信息考虑为“由到达的依赖”，本步骤旨在建立日志条目间的依赖关系。

遍历日志条目列表*R*中的日志条目，可以建立日志条目的有向图*G*，设*G*为

其中*E*和*V*分别为有向图*G*的边和点的集合。*E*中单条边对应每条日志条目， 容易得到具有与相类似的表示，如下

此处的即为有向图*G*中单条有向边的源节点和目的节点，因而*V*集合即为所有组成的节点集合。

通过上述步骤，我们得到了关于日志条目列表*R*的有向图*G*，原有日志文件以行为单位的新改动也将被实时加入到条目库R，系统基于此类新条目增量地更新图G。

（3）攻击链路匹配

本步骤旨在对有向图*G*中每条有向边执行具体分析，匹配得到可能作为攻击链路的异常事件，以待进一步处理。

根据安全报告 ，尽管攻击行为的具体表现不同，一个攻击行为应符合以下攻击链（kill-chain），我们将以该链作为模式匹配对象，

**1.以攻击链作为模式匹配对象，建立攻击链的高级阶段表示**

**以攻击链作为模式匹配对象，对有向图G中的每条有向边E进行异常事件匹配**

**2.**

Initial Compromise -> Establish Foothold -> Internal Recon -> Privilege Escalation将攻击链的四个高级阶段分别表示为。

表格

描述已自动生成

我们首先建立部分与杀伤链中四个高级阶段的对应关系，如Untrusted Read对应Initial Compromise，进而可以通过对有向边链路与的匹配，我们可以识别出可能的攻击链路。其中，我们要求后一个高级阶段的时间（即）要在前一个高级阶段的时间之后，同时，我们剔除掉相邻两个步骤的时间差大于设定值的链路，缩小计算范围。将攻击链路表示为

（4）对应关系发现

由于前述发现的可能攻击链路是根据我们预先建立的部分与的对应关系进行匹配，忽略了其它可能的攻击链路，需要对更多的对应关系进行发现。在本步骤中，我们主要根据已有对应关系发现的攻击链路，使用频繁项集算法对更多的与的对应关系进行发现，即寻找满足设定最小支持度的频繁项集。

同时，将发现的与对应关系加入已有的对应关系，重复执行（3）（4）步骤，即攻击链路匹配与对应关系发现，直到对应关系趋于稳定。