

工业控制系统物理层报警量关联算法

王英州¹, 张耀方¹, 徐有方¹, 赵若菡¹, 郭舒畅¹, 刘红梅³, 张永铮⁴, 王佰玲¹, 刘红日^{1,2}

(1. 哈尔滨工业大学(威海)计算机科学与技术学院, 山东 威海 264209;

2. 威海天之卫网络空间安全科技有限公司, 山东 威海 264209;

3. 中国移动通信集团山东有限公司, 山东 济南 250000; 4. 中资网络信息安全科技有限公司, 北京 100000)

摘要: 工业控制系统遭到攻击后, 可能会导致部分工控设备故障, 负责监控设备工作状态的传感器会在物理层产生报警数据。这些传感器变量被称为报警量, 它们反映了设备的运行状况。然而, 报警数据数量巨大, 且相关性难以确定, 同时报警信息的源头也很难追溯。针对上述问题, 提出了一种针对工业控制系统物理层报警量关联的算法。该算法先通过分析报警记录中的报警开始时间和报警结束时间来量化报警量间的关联度, 在关联度的指导下对报警量进行分类关联, 构建可能的全局报警关联结构, 最后利用K2算法对可能的全局报警关联结构进行评估。同时改进了K2算法的评分函数, 从而保证全局报警关联结构的可解释性。实验结果表明, 该算法得到的全局报警关联结构具有较好的可解释性, 更符合真实工艺流程, 有助于分析报警量在物理层的关联关系, 对指导工业控制系统安全响应有着重要的意义。

关键词: 工业控制系统安全; 报警关联; 追溯报警源头; K2算法

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2025013

Physical layer alarm correlation algorithm for industrial control system

WANG Yingzhou¹, ZHANG Yaofang¹, XU Youfang¹, ZHAO Ruohan¹, GUO Shuchang¹,
LIU Hongmei³, ZHANG Yongzheng⁴, WANG Bailong¹, LIU Hongri^{1,2}

1. School of Computer Science and Technology, Harbin Institute of Technology (Weihai), Weihai 264209, China

2. Weihai Cyberguard Technologies Co., Ltd, Weihai 264209, China

3. China Mobile Communications Group Shandong Co., Ltd, Jinan 250000, China

4. China Assets Cybersecurity Technology Co., Ltd, Beijing 100000, China

Abstract: Malfunctions of some industrial control equipment may be caused by attacks on industrial control systems, and alarm data was generated at the physical layer by sensors responsible for monitoring the equipment's working status. These sensor variables, referred to as alarm quantities, reflected the operating conditions of the equipment. However, the volume of alarm data was massive, and its correlation was difficult to determine. Simulta-

收稿日期: 2024-04-30; 修回日期: 2024-09-28

通信作者: 刘红日, liuhr@hit.edu.cn

基金项目: 国家重点研发计划 (2021YFB2012400)

Foundation Item: The National Key R&D Program of China (2021YBF2012400)

引用格式: 王英州, 张耀方, 徐有方, 等. 工业控制系统物理层报警量关联算法[J]. 网络与信息安全学报, 2025, 11(1): 178-188.

Citation Format: WANG Y Z, ZHANG Y F, XU Y F, et al. Physical layer alarm correlation algorithm for industrial control system[J]. Chinese Journal of Network and Information Security, 2025, 11(1): 178-188.

neously, tracing the source of the alarm data was also challenging. To address these issues, an algorithm for physical layer alarm correlation in industrial control systems was proposed. The correlation between alarm variables was quantified by analyzing the alarm start time and alarm end time in alarm records. Alarm variables were classified and correlated under the guidance of relevance, and potential global alarm correlation structures were constructed. Finally, the potential global alarm correlation structure was evaluated using the K2 algorithm. Additionally, the scoring function of the K2 algorithm was improved to ensure the interpretability of the global alarm correlation structure. Experimental results demonstrate that the global alarm correlation structure obtained by this algorithm exhibits good interpretability and is more consistent with the real process flow. This structure is helpful in analyzing the correlation between alarm variables at the physical layer and is of significant importance in guiding the security response of industrial control systems.

Keywords: industrial control system security, alarm correlation, tracing the source of alarm, K2 algorithm

0 引言

工业控制系统(industrial control system, ICS)是由各种自动化控制组件以及负责采集、监测实时数据的过程控制组件共同构成的业务流程管理系统^[1-2]。随着工业4.0时代的到来, ICS愈加智能化, 同时针对ICS的攻击日益增多, ICS的安全性面临巨大挑战。ICS被攻击可能导致工控设备工作异常^[3], 此时监控工控设备工作状态的传感器会发出报警。如果可以通过分析传感器报警数据找到报警的源头, 就能尽量减少甚至避免资产的损失。然而, 为了保证ICS的可靠性和安全性, 通常不会公开物理层的全局设备连接关系, 因此很难直接利用设备连接关系排查故障源头并解决问题。设计出一种仅通过分析物理层报警数据就能将报警量进行关联的算法非常重要。目前, 针对ICS物理层报警量关联的方法主要有基于模式匹配的方法、基于专家经验的方法和基于因果关系的方法3类。

在基于模式匹配的方法中, Zhou等^[4]基于CloFAST算法提出了一种改进的相似报警序列挖掘方法, 并设计了一种模式蒸馏策略用来合并相似的报警序列并导出紧凑的报警序列模式, 再结合专家知识分析实现了定位报警根源。Hu等^[5]通过设置报警洪泛的识别和提取规则, 定义频繁的报警序列来发现相关模式。Lai等^[6]提出了一种改进的报警序列对齐算法, 通过结合警报的优先级信息和时间戳实现报警序列间的相似性计算, 有助于报警数据分类管理和报警序列因果分析。Niyazamand等^[7]改进了PrefixSpan算法来提取报警序列中的频繁模式, 可以更加准确地识别出相似

的报警序列, 将报警序列进行关联。Li等^[8]针对报警数据量大、计算效率较低的问题, 提出了区块化处理报警信息方法。该方法将报警数据转化为报警区块和非报警区块的时间节点信息, 用两个块间的匹配度作为相似度, 在保证信息完整的情况下减少了计算量。鲁显光等^[9]提出了一种基于改进FP Grown的警报关联算法, 在FP Grown的基础上引入二维表和频繁项挖掘方法, 能够在较好地处理稀疏数据的同时, 对原始算法的内存消耗和运行时间进行较好的优化。

在基于专家经验的方法中, Rodrigo等^[10]结合报警日志、过程数据和过程连接关系信息来分析报警泛洪, 并对报警进行分组, 将源自相同过程异常的报警变量进行隔离, 确定报警洪泛中的报警关联, 识别报警根源。Simeu等^[11]考虑报警集合冗余的同时, 分析识别出具体报警的特征, 再利用动态故障树来定位报警根源。Guo等^[12]通过对报警时间戳的分析和挖掘, 完善了报警数据处理过程, 并利用时间约束网络实现报警源定位。Thambirajah等^[13]从过程平面图获取生产过程连通性的基本信息, 并根据测量数据得到因果矩阵, 通过将两者结合起来寻找报警的传播路径, 从而定位报警源。

在基于因果关系的方法中, Liu等^[14]采用格兰杰因果关系分析法对工业流程中的不同功能子模块进行分析, 并利用最大生成树算法得到每个功能子模块的有向无环图, 最后通过子图拼接得到整个工业流程的报警量因果关联图。Chen等^[15]对格兰杰因果关系分析法进行了改进, 使用多元条件格兰杰因果来构建工业报警量间的因

果关系图。改进后的关联算法可以将时间粒度细化到工业流程中的每一个步骤,而且得到的因果关系图具有更好的可解释性。Hu 等^[16]采用传递熵方法来分析报警量关联关系,建立因果关系图,再对因果关系进行简化,最终通过分析得到报警量关联结构。Shu 等^[17]对传递熵公式进行了改进,用相邻的报警状态来加强报警量的因果关系,提高了因果关系图的可解释性。Su 等^[18]用互信息来简化因果关系图,利用条件互信息设置因果关系阈值,完成因果关系的简化。Abreu 等^[19]先通过传递熵计算出报警量关联关系,再引入 K2 算法实现报警量关联关系的简化。Wang 等^[20]通过建立报警变量的贝叶斯网络对产生报警的根本原因进行分析。

分析国内外的研究工作,发现目前面向 ICS 的报警量关联方法存在如下问题:基于模式匹配的方法对报警数据的完整性和连续性提出了较高的要求,当某些字段缺省时,会导致匹配精度大幅下降,而且对报警序列特征的数值化计算也可能导致数据本身部分信息丢失。基于专家经验的方法需要大量的工业过程知识,然而真实工控场景中业务流程复杂、报警点位多,梳理过程连接等信息较为困难,而且可迁移性差,更换新的工控场景后,通常需要重新标注数据,消耗的人力成本高。相较前两种方法,基于因果分析的方法对报警数据的要求较低,而且可以不借助专家经验值,仅从报警数据本身的关系进行挖掘分析,泛用性较强。因此,本文选择改进现有的因果分析方法。因果分析法中的格兰杰因果分析仅能分析两个报警量间是否存在线性关系,而 ICS 中报警量间通常为非线性关系,传递熵虽然可以用于分析非线性关系,但传递熵需要构造的状态空间数呈指数级,计算的时间复杂度过高,难以应用于拓扑较复杂的 ICS 场景。

针对上述问题,本文提出了一种适用于 ICS 物理层报警量关联的算法,首先通过分析报警量的报警记录找到报警量间的关联度,再根据关联度的强弱对报警量进行分类关联,根据关联结果能构造出可能的全局报警量关联结构,最后利用 K2 算法对可能的全局报警量关联结构进行评分,评分最高的关联结构为当前评估算法下找到的最优解。

1 报警量关联前置知识

ICS 被广泛应用于电力、水处理、石油和天然气、化工、交通等领域,其核心组件包括数据采集与监控系统、分布式控制系统、可编程逻辑控制器(PLC)、可编程自动化控制器、远程终端单元、传感器等。ICS 的架构及安全风险示意如图 1 所示,ICS 架构由企业管理层、生产管理层、过程监控层、现场控制层和现场设备层 5 层组成。

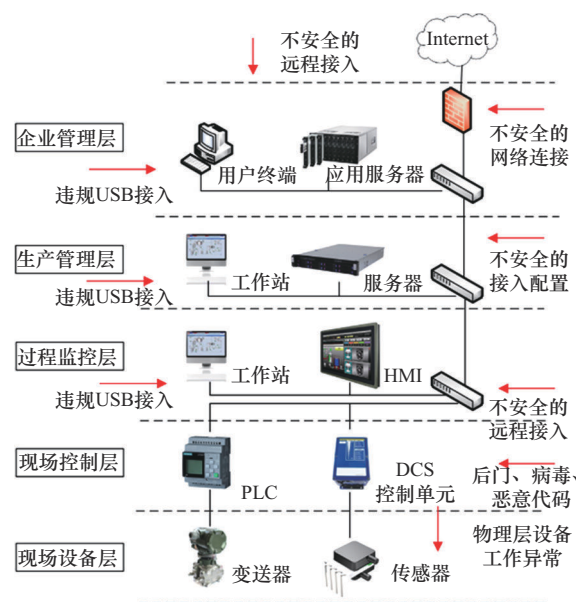


图 1 ICS 的架构及安全风险示意

Figure 1 Schematic of ICS hierarchy and security risks

企业管理层可为企业决策层人员提供决策运行手段;生产管理层负责对生产过程进行管理;过程监控层对生产过程数据进行采集和监控,并利用人机界面实现人机交互;现场控制层主要包括可编程逻辑控制器、分布式控制系统(DCS)等,负责对各执行设备进行控制;现场设备层对生产过程进行感知与操作,主要包括各类过程传感设备与执行设备单元^[21]。ICS 遭受攻击,可能导致部分工控设备处于异常状态,对应传感器获取的采样值一旦偏离正常值,就会触发报警。本文通过收集现场设备层的报警数据来分析出报警量间的关联结构。

为了方便描述 ICS 物理层报警量关联算法,下面先给出相关定义以及符号表示。

定义 1 传感器报警模式

ICS 中通常部署了大量的传感器,用来监控

工控设备的工作状态。当工控设备的工作状态出现异常时，传感器获取的工控设备采样值就会偏离正常值，从而触发报警。规定“0”表示正常工作状态，“1”表示报警状态。当采样值小于低阈值时，称为低报警；当采样值大于高阈值时，称为高报警。

定义2 传感器报警记录

传感器报警记录记录的是一段时间内传感器的工作状态。这是因为传感器一旦触发报警，就会在一段连续时间内都处于报警状态。报警记录可以用二元组 (Start, End) 表示。其中，Start 表示报警开始时间，End 表示报警结束（即设备恢复正常状态）的时间。

图2所示为传感器1和传感器2在一段时间内的报警记录，传感器1在这段时间内共有两条报警记录，分别为 (1, 3) 和 (5, 7)；传感器2在这段时间内也有两条报警记录，分别为 (2, 4) 和 (6, 8)。

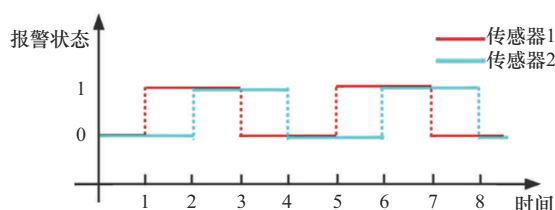


图2 传感器1和传感器2的报警记录
Figure2 Alarm records for sensor 1 and sensor 2

定义3 报警时间相关性

报警时间相关性指通过分析两个传感器报警记录的重叠程度来量化报警量间的关联关系。若两个传感器的报警记录中的报警时间存在重叠部分，即存在同时报警的情况，则认为这两个报警量存在关联。可以利用式(1)对报警量间关联关系进行量化。

$$ATC(A_i, A_j) = \frac{\sum_{AR_{iu} \cap AR_{jv} \neq \emptyset} \text{OverlapTime}(AR_{iu}, AR_{jv})}{\sum_{AR_{iu} \cap AR_{jv} \neq \emptyset} \max(End_{jv} - Start_{iu}, End_{iu} - Start_{jv})} \quad (1)$$

其中， A_i 和 A_j 分别表示报警量 i 和报警量 j ， AR 表示报警记录， AR_{iu} 和 AR_{jv} 分别表示报警量 i 的第 u 条报警记录和报警量 j 的第 v 条报警记录，

$\text{OverlapTime}(AR_{iu}, AR_{jv})$ 表示报警量 i 的第 u 条报警记录和报警量 j 的第 v 条报警记录的重叠报警时间。式(1)的分子相当于统计了两个报警量的重叠报警时间之和。图2中两个报警量在 $t=2$ 到 $t=3$ 以及 $t=6$ 到 $t=7$ 这两段时间均处于报警状态，因此式(1)的分子等于两段报警重叠时间之和，为2。Start 和 End 分别对应报警量的报警记录中报警开始时间和报警结束时间。式(1)的分母表示两个报警量总的报警持续时间。ATC 值越大，表示两个报警量之间的关联性越强。

定义4 报警频率相关性

报警频率相关性指通过统计两个报警量报警时间的重叠次数占总报警次数的比例来量化报警量间的关联关系。两个报警量间的报警频率相关性计算式如下。

$$AFC(A_i, A_j) = \frac{T}{C(A_i) + C(A_j) - T} \quad (2)$$

式(2)中 T 表示报警记录中报警时间出现重叠的次数， C 表示对报警记录进行计数， $C(A_i)$ 表示报警量 i 在时间 t 内报警记录的数量， $C(A_j)$ 表示报警量 j 在时间 t 内报警记录的数量。图2中两个报警量出现了两次共同报警的状态，因此 T 取2，由于两个报警量在时间 t 内各有两条报警记录，因此 $C(A_i)$ 和 $C(A_j)$ 均取2。AFC 值越大，表示两个报警量之间的关联性越强。

定义5 报警量关联度

报警量关联度 R 结合了报警时间相关性和报警频率相关性，对报警量间的关联性强弱进行综合量化。报警时间相关性越强，报警频率相关性越强，则报警量关联度越大。报警量关联度计算式如下。

$$R(A_i, A_j) = ATC(A_i, A_j) \times AFC(A_i, A_j) \quad (3)$$

2 算法描述

ICS 物理层报警量关联算法分为3个步骤：计算报警量间的关联矩阵；根据报警间关联性的强弱对报警量进行分类；利用 K2 算法对可能的报警关联结构进行评估。下面对这3个步骤分别进行阐述。

2.1 计算报警量间的关联矩阵

对于任意两个报警量 A_i 和 A_j ，统计这两个报警量在报警记录中的报警重叠时间和报警重叠次

数, 计算报警量时间相关性 ATC 和报警量频率相关性 AFC, 将 ATC 和 AFC 相乘就可以得到任意两个报警量间的关联度 R 。计算报警量间的关联度的算法伪代码如下。

算法 1 报警量相关性量化算法

```

输入 报警量  $A_i$  的报警记录 Alarm_record $i$ ;
报警量  $A_j$  的报警记录 Alarm_record $j$ 
输出 报警量  $A_i$  和  $A_j$  的关联度
1)  $C(A_i), C(A_j) = \text{len}(\text{Alarm\_record}_i), \text{len}(\text{Alarm\_record}_j)$ 
2)  $\Delta t, \text{all\_time} = 0, 0$ 
3) for each Start $i$ , End $i$   $\in$  Alarm_record $i$  do
4)   for each Start $j$ , End $j$   $\in$  Alarm_record $j$  do
5)     Time $i$   $\leftarrow$  End $i$  - Start $i$ 
6)     Time $j$   $\leftarrow$  End $j$  - Start $j$ 
7)     if max(End $i$  - Start $j$ , End $j$  - Start $i$ ) > Time $i$  + Time $j$  then
8)       Skip
9)     else
10)      if Start $i$  < End $j$  < Start $i$  or Start $j$  < End $i$  < Start $j$  then
11)         $\Delta t += \min(\text{End}_i - \text{Start}_j, \text{End}_j - \text{Start}_i)$ 
12)         $\text{all\_time} += \max(\text{End}_i - \text{Start}_j, \text{End}_j - \text{Start}_i)$ 
13)      else
14)         $\Delta t += \min(\text{Time}_i, \text{Time}_j)$ 
15)         $\text{all\_time} += \max(\text{Time}_i, \text{Time}_j)$ 
16)      end if
17)    end if
18)  end for
19) end for
20)  $\text{ATC}(A_i, A_j) \leftarrow \Delta t / \text{all\_time}$ 
21)  $\text{AFC}(A_i, A_j) \leftarrow T / (C(A_i) + C(A_j) + T)$ 
22)  $R \leftarrow \text{ATC}(A_i, A_j) * \text{AFC}(A_i, A_j)$ 
23) return  $R$ 

```

算法 1 通过比较两个报警量总的报警持续时间和报警量分别报警持续时间之和的大小关系, 判断两个报警量是否存在同时报警的状态, 更新报警重叠时间和报警重叠次数计数器, 完成报警量的报警时间相关性和报警频率相关性的计算。最后这两个结果相乘即报警量的关联度 R 。

2.2 报警量分类

在使用 K2 算法确定最合理的警报关联结构前需要确定报警量的输入顺序。若报警量个数为 n , 则所有报警量输入顺序的组合为 $n!$, 远超多项式时间, 因此需要先确定报警量的大致排列顺序。通过算法 1 可以获取任意两个报警量间的关联度, 将结果存入矩阵中得到报警量关联度矩阵。利用两个报警量间的关联度 R 强弱程度可以进行类别的划分, 关联程度相近的报警量可以放到一类中, 再按照关联度从高到低的顺序将类和类进行拼接, 就可以大幅减少报警量的组合可能。报警量关联得到类的方法如下。

1) 首先根据报警量相关性算法可以获得报警量间的关联度矩阵, 对于报警量集合 A 中尚未分类的报警量, 计算其他所有未分类报警量的均值 t 作为当前阈值。当报警量关联度 R 大于阈值 t 时, 将报警量加入类中。

2) 更新未分类报警量集合 A 和阈值 t , 将剩余报警量中大于阈值的进行分类关联。

3) 重复上述操作, 直到报警变量集合 A 为空, 即所有报警量都已经分类。

2.3 关联结构计算

节点关联网络结构是一种有向图模型, 表示一组随机变量之间的因果关系及其条件依赖性。K2 算法是经典的基于评分搜索的节点关联网络结构算法, 该算法的主要思想是: 在一个无环的有向图中, 规定每个节点的最大父节点数量后, 输入节点序, K2 算法在处理每一个节点时, 会搜索该节点的父节点集, 使得该节点与其父节点所在的局部网络评分值不断增加, 当达到设定的最大父节点数量或者评分值收敛不再增加时, 算法终止, 输出评分值最高的网络结构。

定义 D 为给定的节点序组合数据集, G 表示一种节点关联结构。K2 算法的本质就是在给定数据集 D 的条件下找到评分结构最高的节点关联结构 G , 即求条件概率 $p(G|D)$ 的最大值, 利用式(4)可以对条件概率进行化简。

$$p(G|D) = \frac{p(G, D)}{p(D)} \quad (4)$$

由于 $p(D)$ 与节点关联结构 G 无关, 仅与自身数据样本分布有关, 因此 $p(D)$ 的取值为常数, 则式(4)中等号左边的条件概率 $p(G|D)$ 和右边的联

合概率 $p(G|D)$ 变化趋势一致。当 $p(G|D)$ 取得最大值时, $p(G|D)$ 也取得最大值。K2算法通过评分函数量化 $p(G|D)$ 的合理性, 评分函数的计算式如下。

$$\max(p(G, D)) = c \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (5)$$

式(5)中常数 c 表示 $p(D)$ 。从整体来看, 式(5)对每个节点和其父节点取值对应关系的概率结果进行累乘, 再将不同节点的计算结果相乘得到评分。 r_i 表示节点 x_i 的所有取值的个数, q_i 表示节点 x_i 所有父节点的取值个数, N_{ij} 表示节点 x_i 的父节点取值为 j 的样本数量, N_{ijk} 表示节点 x_i 的取值为 k 且 x_i 的父节点取值为 j 的样本数量。 N_{ij} 的取值等于将节点 x_i 父节点取值为 j 时, x_i 的所有可能取值的样本个数累加, 计算式如下。

$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk} \quad (6)$$

为了简化计算, 全局网络结构可视为可拆分结构, 则全局网络结构的评分可以用若干局部网络结构的评分相乘得到。局部网络结构的评分彼此独立, 互不影响。因此, 可以利用贪心策略, 当所有的局部网络结构评分取得最大值时, 全局网络结构评分也取得最大值。式(7)为局部网络结构的计算式。

$$g(x_i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (7)$$

$$G'(G, D) = \prod_{i=1}^n g(x_i, \pi_i) \quad (8)$$

式(7)中, π_i 表示节点 x_i 的父节点集, 根据贪心策略, 只需要找到 $g(x_i, \pi_i)$ 的最大值, 即可最大化 $G'(G, D)$ 。

K2算法规定每个节点的父节点数量上限, 对于节点序中的每一个节点 x_i , 将排在它之前的 $i-1$ 个节点接入节点 x_i 的候选父节点集合。如果候选父节点 x_j 对局部结构的评分有提升, 则将节点 x_j 加入 x_i 的父节点集 π_i 中。重复上述过程, 直到网络结构的评分不再增加, 或达到 x_i 的父节点数量上限时, K2算法结束, 输出最优节点关联结构 G 。K2算法的伪代码如下。

算法2 K2算法

输入 数据集 D ; 节点集 A ; 候选父节点集

合Candidate_fa; 父节点上限 m

输出 最优节点关联结构 G

```

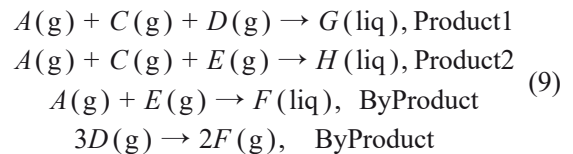
1) for node  $x_i \in A$  do
2)    $\pi_i \leftarrow \varphi$ 
3)    $P_{old} \leftarrow \text{Score}(i, \pi_i)$ 
4)   flag  $\leftarrow \text{True}$ 
5)   While flag and  $\text{len}(\pi_i) < m$  do
6)     for node  $x_j \in \text{Candidate\_fa}$  do
7)        $P_{new} \leftarrow \text{Score}(i, \pi_i \cup \{x_j\})$ 
8)       if  $P_{new} > P_{old}$  then
9)          $P_{old} \leftarrow P_{new}$ 
10)         $\pi_i \leftarrow \pi_i \cup \{x_j\}$ 
11)      end if
12)    if 已经遍历Candidate_fa集合的所有元素 then
13)      flag  $\leftarrow \text{False}$ 
14)    end if
15)  end for
16) end while
17) end for
18)  $G \leftarrow \bigcup \{\pi_k\}$  for  $k$  in range( $\text{len}(A)$ )
19) return  $G$ 

```

3 实验

3.1 数据集介绍

使用文献[19]公开的工业报警数据集进行实验验证, 该数据集从田纳西-伊斯曼(Tennessee-Eastman)仿真平台^[22]上获取, 得到的数据集具有时变、强耦合和非线性特征。田纳西-伊斯曼仿真平台模拟了一套控制乙烯生产的ICS^[23-24], 该过程包括 A 、 B 、 C 、 D 、 E 、 F 、 G 、 H 共8部分产物, 反应表达式如下。



田纳西-伊斯曼仿真平台模拟的ICS共包含12个控制变量和41个过程测量变量。文献[19]公开的数据集选择了41个过程测量变量中的8个测量变量进行重点监测, 同时记录8个过程测量变量的工作状态和采样值, 利用设置的高低阈值可以将每一个测量变量划分为高报警状态和低报警

状态。因此, 8 个过程测量变量可以得到 16 个报警状态, 经过 760 h 的仿真模拟获得约 70 000 条报警数据, 数据采样频率约为 40 s。变量关系如表 1 所示。

表 1 变量关系
Table 1 Variable relationship

过程变量	描述	单位
x_1	A 进料 (流 1)	km ³ /h
x_2	D 进料 (流 2)	kg/h
x_3	E 进料 (流 3)	kg/h
x_6	反应器进料速度 (流 6)	km ³ /h
x_7	反应器压力	kPa
x_8	反应器等级	%
x_9	反应器温度	°C
x_{21}	反应器冷却水温度	°C

3.2 数据处理

对报警数据进行预处理, 删除未发生报警的数据。利用算法 1 可以得到报警量关联度矩阵, 关联度矩阵见表 2。表 2 的行和列均表示报警量, 如 x_{1_H} 表示测量变量 x_1 高报警, x_{1_L} 表示测量变量 x_1 低报警。

根据报警量关联矩阵和第 2.2 节提出的报警量关联方法, 维护未分类的报警量集合 A 和分类阈值 t , 每次将关联度 R 大于分类阈值 t 的报警量加入一个新的类中, 同时从集合 A 移除这些报警量, 更新分类阈值 t , 重复上述过程, 可以将所有报警量进行关联分类。报警量关联分类结果如表 3 所示。

表 2 报警量关联度矩阵
Table 2 Alarm quantity correlation matrix

	x_{1_H}	x_{7_H}	x_{8_H}	x_{9_H}	x_{21_H}	x_{1_L}	x_{6_L}	x_{7_L}	x_{21_L}
x_{1_H}	0	100×10^{-5}	6.6×10^{-5}	20×10^{-5}	50×10^{-5}	0	7.3×10^{-5}	100×10^{-5}	80×10^{-5}
x_{7_H}	100×10^{-5}	0	30×10^{-5}	130×10^{-5}	40×10^{-5}	0	20×10^{-5}	0	80×10^{-5}
x_{8_H}	6.6×10^{-5}	30×10^{-5}	0	220×10^{-5}	10×10^{-5}	0	0	0	10×10^{-5}
x_{9_H}	20×10^{-5}	130×10^{-5}	220×10^{-5}	0	5.5×10^{-5}	0	0	0	100×10^{-5}
x_{21_H}	50×10^{-5}	40×10^{-5}	10×10^{-5}	5.5×10^{-5}	0	280×10^{-5}	300×10^{-5}	450×10^{-5}	0
x_{1_L}	0	0	0	0	280×10^{-5}	0	200×10^{-5}	290×10^{-5}	0
x_{6_L}	7.3×10^{-5}	20×10^{-5}	0	0	300×10^{-5}	200×10^{-5}	0	310×10^{-5}	2.2×10^{-5}
x_{7_L}	100×10^{-5}	0	0	0	450×10^{-5}	300×10^{-5}	310×10^{-5}	0	0
x_{21_L}	80×10^{-5}	80×10^{-5}	10×10^{-5}	100×10^{-5}	0	0	2.2×10^{-5}	0	0

表 3 报警量关联分类结果
Table 3 Alarm quantity correlation
classification results

类别	报警变量	阈值 t
1	x_{1_L} , x_{7_L} , x_{21_H} , x_{6_L}	0.002524
2	x_{8_H} , x_{9_H}	0.001517
3	x_{1_H} , x_{7_H}	0.001105
4	x_{21_L}	0.000955

报警量共分为 4 类, 4 个类别中的报警量按照首次报警时间从早到晚的顺序排列。在排列组合所有可能的节点顺序时, 对于每个类来说, 固定第一个报警变量, 即首次报警的报警量, 将类中剩余报警量进行全排列组合; 对于不同类来说, 按照阈值 t 从大到小的顺序将这 4 个类进行拼接。最终可以得到 K2 算法所有可能的报警量输入节点序。输入节点序如表 4 所示。

将这 6 种可能的报警量节点序作为 K2 算法的输入, K2 算法通过分析节点序可以获得每个报警量节点的候选父节点集合, 利用式(7)和式(8)计算评分函数以决定是否选择候选父节点集合中的元素作为最终结果。在最多父节点个数的约束下, K2 算法能计算当前节点序的父节点集合以及报警量间的关联结构, 6 种根据报警量节点序生成的报警量关联结构如图 3 所示。图 4 给出了最优关联结构 (关联结构 1), 节点序为 x_{1_L} , x_{21_H} , x_{6_L} , x_{7_L} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_H} , x_{21_L} 。

表 4 输入节点序

Table4 Input node sequence

序号	输入节点序
1	x_{1_L} , x_{21_H} , x_{7_L} , x_{6_L} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_H} , x_{21_L}
2	x_{1_L} , x_{21_H} , x_{6_L} , x_{7_L} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_H} , x_{21_L}
3	x_{1_L} , x_{7_H} , x_{21_H} , x_{6_L} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_L} , x_{21_L}
4	x_{1_L} , x_{7_H} , x_{6_L} , x_{21_H} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_H} , x_{21_L}
5	x_{1_L} , x_{6_H} , x_{21_H} , x_{7_L} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_H} , x_{21_L}
6	x_{1_L} , x_{6_H} , x_{7_L} , x_{21_H} , x_{8_H} , x_{9_H} , x_{1_H} , x_{7_H} , x_{21_L}

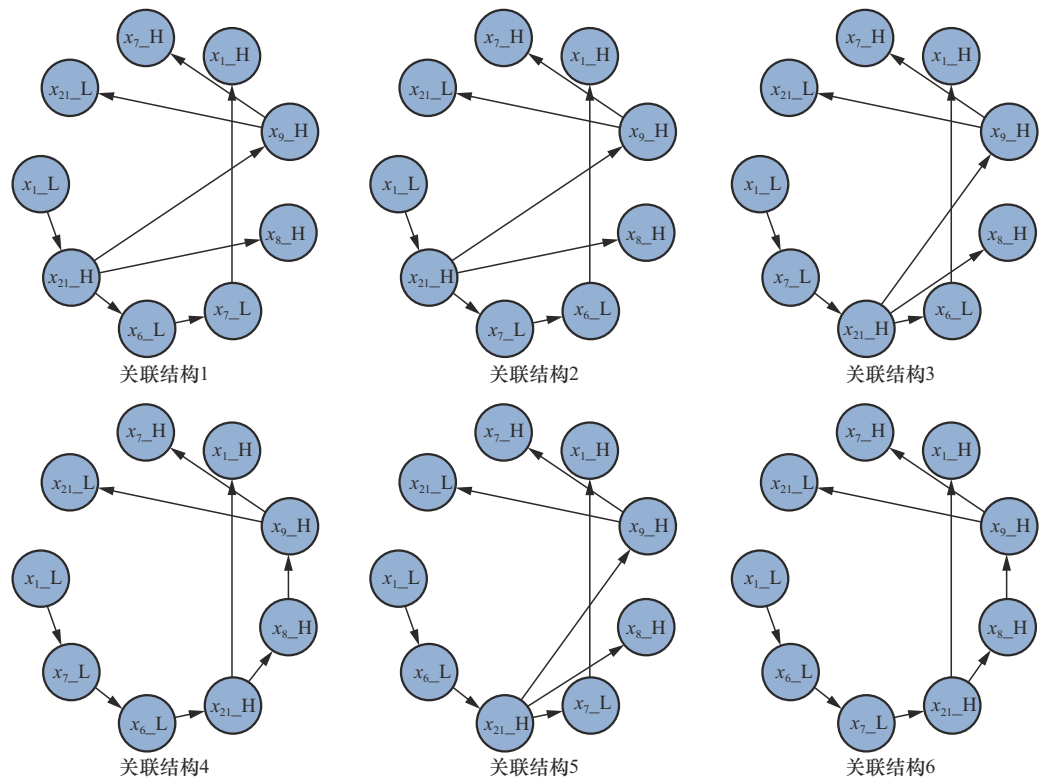


图 3 根据报警量节点序生成的报警量关联结构

Figure 3 The alarm quantity correlation structure generated based on the order of alarm quantity nodes

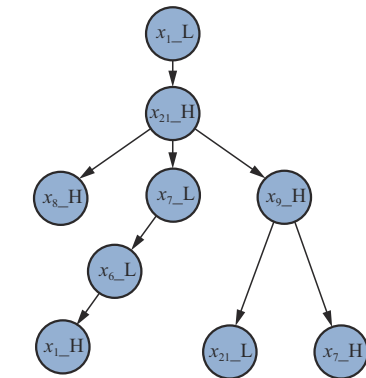


图 4 本文方法推算出的最优关联结构

Figure 4 The optimal correlation structure is deduced by the proposed method

3.3 实验结果及分析

图 3 中关联结构 1 的评分最高，表示这是最合理的报警量关联结构，关联结构 1 的根节点为 x_{1_L} ，第一条关联关系为 $x_{1_L} \rightarrow x_{21_H}$ ，表示 A 进料的亏损导致反应式(9)的平衡被打破，进而导致反应器冷却水温度升高；报警量 x_{21_H} 指向了 3 个报警量 x_{6_L} 、 x_{8_H} 、 x_{9_H} ，分别表示反应器冷却水温度升高导致反应器进料速度降低、反应器等级升高和反应器温度升高； $x_{6_L} \rightarrow x_{7_L}$ 表示反应器进料速度减慢导致反应器压力降低； $x_{7_L} \rightarrow x_{1_H}$ 表示反应器压力降低

导致进料 A 增加（这种情况是控制器导致的，控制器检测到进料 A 在反应过程中占比较少，因此对 A 进料进行补充）；剩下的关联关系为 $x_{9_H} \rightarrow x_{7_H}$ 和 $x_{9_H} \rightarrow x_{21_L}$ ，分别表示反应器温度升高导致反应器压力升高和反应器冷却水温度降低。

下面通过分析其他 5 种关联结构的不合理性来证明 K2 算法对报警量关联结构评估的准确性。关联结构 2 和关联结构 1 的区别在于 x_{6_L} 和 x_{7_L} 的顺序。从表 1 中可以得知 x_6 表示反应器的进料速度， x_7 表示反应器压力。从工艺流程的角度分析，反应器进料速度会影响反应器的压力，而关联结构 2 认为 $x_{7_L} \rightarrow x_{6_L}$ ，关联结果和工业流程相违背，缺乏合理性。关联结构 3 和关联结构 4 也存在类似的问题， x_{7_L} 都位于 x_{6_L} 的上游位置。关联结构 5 和关联结构 1 的区别在于 x_{6_L} 和 x_{21_H} 的顺序。由表 1 可知， x_6 表示反应器进料速度， x_{21} 表示反应器冷却水温度。从工艺流程的角度分析，高温可能会导致反应器进料速度降低， x_{21_H} 应该处于 x_{6_L} 的上方。关联结构 6 中 x_{21_H} 和 x_{9_H} 没有建立直接的关联关系。从表 1 中可知， x_{21} 和 x_9 都是反应器附近的温度参数变量。将温度变量关联到一起更符合工艺流程。综上所述，其他 5 种关联结构存在一些不合理的连接关系，关联结构 1 能够较好地满足报警量在工艺流程中的关联关系，因此 K2 算法评估关联结构的效果很好。

除了能够找到当前评估算法下的最优关联结构，还可以追溯报警源头。当发现某个传感器发生故障时，可以参考图 4 的关联结构逐级向上排查故障，直到找到引发故障的源头。

本文通过报警量相关性计算获得报警量间的节点序排列，再根据 K2 算法对节点序进行评分，找出评分最高的报警量关联结构。文献[19]和本文算法类似，都是通过确定 K2 算法的输入节点序来获得最优关联结构，故将该算法与本文算法进行对比，文献[19]推算出的最优关联结构如图 5 所示，本文算法和文献[19]推算出的报警量关联关系结果对比如表 5 所示。

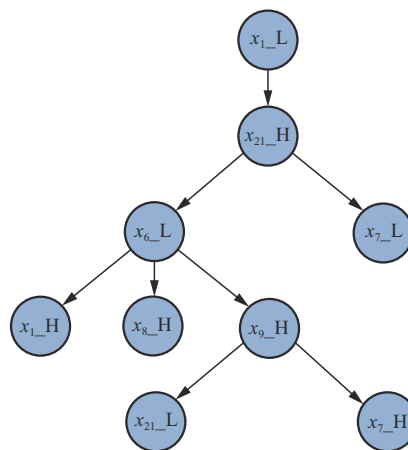


图 5 文献[19]推算出的最优关联结构

Figure 5 The optimal correlation structure is deduced in literature[19]

表 5 报警量关联关系结果对比

Table 5 Comparison of alarm quantity correlation results

报警量关联关系	本文算法	文献 [19]
$x_{1_L} \rightarrow x_{21_H}$	√	√
$x_{21_H} \rightarrow x_{6_L}$	√	√
$x_{21_H} \rightarrow x_{7_L}$		√
$x_{6_L} \rightarrow x_{7_L}$	√	
$x_{6_L} \rightarrow x_{8_H}$		√
$x_{6_L} \rightarrow x_{1_H}$		√
$x_{6_L} \rightarrow x_{9_H}$		√
$x_{7_L} \rightarrow x_{1_H}$	√	
$x_{21_H} \rightarrow x_{9_H}$	√	
$x_{21_H} \rightarrow x_{8_H}$	√	
$x_{9_H} \rightarrow x_{7_H}$	√	√
$x_{9_H} \rightarrow x_{21_L}$	√	√

本文算法和文献[19]得到的报警量关联结构的区别主要在于 x_{6_L} 、 x_{7_L} 、 x_{8_H} 和 x_{9_H} 的连接关系。从变量本身的意义来看，这几个报警量都与反应器参数相关。文献[19]的结果中存在一条不合理的关联关系 $x_{21_H} \rightarrow x_{7_L}$ (x_{21} 表示反应器冷却水温度， x_7 表示反应器的压力)。本文方法改善了这个问题，将 x_{7_L} 和 x_{6_L} 进行关联， x_6 表示反应器的进料速度，和报警量 x_7 的关联性较强，因此本文推算出的关联关系 $x_{6_L} \rightarrow x_{7_L}$ 较合理，更贴合工业流程逻辑。

4 结束语

本文对报警量间的关联因素进行了深入研究和分析,并结合K2算法对报警量关联结构进行评估,提出了一种针对ICS物理层的报警关联算法。该算法利用报警时间相关性和报警频率相关性对报警量关联度进行量化,再根据关联度量化结果进行分类,从而大幅减少报警节点序的可能,减少K2算法的运算压力。将K2算法的输入预处理时间从阶乘级别大幅度降低。实验结果证明,本文算法具有较强的可解释性,利用算法分析出的关联结构能够较好地贴合工业流程。在攻击手段越来越复杂的今天,仅通过报警数据分析出报警量间的关联关系无疑对ICS的安全防护有着重要的意义。下一步将继续优化报警量关联度计算方法,尽可能增加关联度的数量级,提高鲁棒性,进一步提高计算的准确度。

参考文献:

- [1] 杨婷, 张嘉元, 黄在起, 等. 工业控制系统安全综述[J]. 计算机研究与发展, 2022, 59(5): 1035-1053.
YANG T, ZHANG J Y, HUANG Z Q, et al. Survey of industrial control systems security[J]. Journal of Computer Research and Development, 2022, 59(5): 1035-1053.
- [2] 黄涛, 王郅伟, 刘家池, 等. 工控协议安全研究综述[J]. 通信学报, 2024, 45(6): 60-74.
HUANG T, WANG Z W, LIU J C, et al. Survey on industrial control protocol security research[J]. Journal on Communications, 2024, 45(6): 60-74.
- [3] 唐士杰, 袁方, 李俊, 等. 工业控制系统关键组件安全风险综述[J]. 网络与信息安全学报, 2022, 8(3): 1-17.
TANG S J, YUAN F, LI J, et al. Review on security risks of key components in industrial control system[J]. Chinese Journal of Network and Information Security, 2022, 8(3): 1-17.
- [4] ZHOU B Y, HU W K, CHEN T W. Pattern extraction from industrial alarm flood sequences by a modified CloFAST algorithm[J]. IEEE Transactions on Industrial Informatics, 2022, 18(1): 288-296.
- [5] HU W K, CHEN T W, SHAH S L. Detection of frequent alarm patterns in industrial alarm floods using itemset mining methods[J]. IEEE Transactions on Industrial Electronics, 2018, 65(9): 7290-7300.
- [6] LAI S Q, YANG F, CHEN T W. Accelerated multiple alarm flood sequence alignment for abnormality pattern mining[J]. Journal of Process Control, 2019, 85(C): 44-57.
- [7] NIYAZMAND T, IZADI I. Pattern mining in alarm flood sequences using a modified PrefixSpan algorithm[J]. ISA Transactions, 2019, 90: 287-293.
- [8] 李俊杰, 杨博, 李宏光. 基于报警时序信息挖掘的报警关联分析方法[J]. 信息与控制, 2019, 48(1): 29-34.
- LI J J, YANG B, LI H G. Alarm correlation analysis approach based on alarm time information mining[J]. Information and Control, 2019, 48(1): 29-34.
- [9] 鲁显光, 杜学绘, 王文娟. 基于改进FP growth的告警关联算法[J]. 计算机科学, 2019, 46(8): 64-70.
LU X G, DU X H, WANG W J. Alert Correlation Algorithm Based on Improved FP Grown[J]. Computer Science, 2019, 46(8): 64-70.
- [10] RODRIGO V, CHIOUA M, HAGGLUND T, et al. Causal analysis for alarm flood reduction[J]. IFAC PapersOnLine, 2016, 49(7): 723-728.
- [11] SIMEU A, LEFEBVREL A, DERAÏN J P. A methodology of alarm filtering using dynamic fault tree[J]. Reliability Engineering & System Safety, 2011, 96(2): 257-266.
- [12] GUO W X, WEN F S, LIAO Z W, et al. An analytic model-based approach for power system alarm processing employing temporal constraint network[J]. IEEE Transactions on Power Delivery, 2010, 25(4): 2435-2447.
- [13] THAMBIRAJAH J, BENEBAAS L, BAUER M, et al. Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history[J]. Computers and Chemical Engineering, 2009, 33(2): 503-512.
- [14] LIU Y, CHEN H S, WU H B, et al. Simplified granger causality map for data-driven root cause diagnosis of process disturbances[J]. Process Control, 2020, 95: 45-54.
- [15] CHEN H S, YAN Z B, ZHANG X L, et al. Root cause diagnosis of process faults using conditional granger causality analysis and maximum spanning tree[J]. IFAC PapersOnLine, 2018, 51(18): 381-386.
- [16] HU W K, CHEN T W, SHAH S L, et al. Cause and effect analysis for decision support in alarm floods[J]. IFAC PapersOnLine, 2017, 50(1): 13940-13945.
- [17] SHU Y D, ZHAO J S. Data-driven causal inference based on a modified transfer entropy[J]. Computers & Chemical Engineering, 2013, 57: 173-180.
- [18] SU J J, WANG D J, ZHANG Y N, et al. Capturing causality for fault diagnosis based on multi-valued alarm series using transfer entropy[J]. Entropy, 2017, 19(12): 663.
- [19] ABREU D R S, NUNES Y T, et al. A method for detecting causal relationships between industrial alarm variables using transfer entropy and K2 algorithm[J]. Process Control, 2021, 106: 142-154.
- [20] WANG J, YANG Z, SU J, et al. Root-cause analysis of occurring alarms in thermal power plants based on Bayesian networks[J]. International Journal of Electrical Power & Energy Systems, 2018, 103: 67-74.
- [21] 吴越, 曹国彦. 面向工业传感网络的时间序列异常检测综述[J]. 网络与信息安全学报, 2024, 10(4): 17-36.
WU Y, CAO G Y. Survey of time series anomaly detection for industrial sensor networks[J]. Chinese Journal of Network and Information Security, 2024, 10(4): 17-36.
- [22] ZHU Q X, MENG Q Q, WANG J P, et al. Novel causal network modeling method integrating process knowledge with modified transfer entropy: a case study of complex chemical processes[J]. Ind. Eng. Chem. Res, 2017, 56: 14282-14289.
- [23] MENEGOZZO G, DALL'ALBA D, FIORINI P. CIPCaD-Bench: continuous industrial process datasets for benchmarking causal dis-

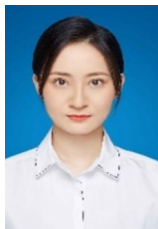
covery methods[C]//Proceedings of IEEE 18th International Conference on Automation Science and Engineering, Piscataway: IEEE Press, 2022: 2124-2131.

- [24] CHEN X L, WANG J, DING S X. Complex system monitoring based on distributed least squares method[J]. IEEE Transactions on Automation Science and Engineering, 2021, 18(4): 1892-1900.

[作者简介]



王英州 (1999-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学(威海)硕士生, 主要研究方向为工业控制系统安全。



张耀方 (1998-), 女, 黑龙江哈尔滨人, 哈尔滨工业大学(威海)博士生, 主要研究方向为工业控制系统攻击路径。



徐有方 (1999-), 男, 河南南阳人, 哈尔滨工业大学(威海)硕士生, 主要方向为工业控制系统安全、网络安全。



赵若菡 (2003-), 女, 山东济宁人, 主要研究方向为工业控制系统安全。



郭舒畅 (2004-), 女, 吉林长春人, 主要研究方向为网络与信息安全。



刘红梅 (1975-), 女, 山东烟台人, 山东大学高级工程师, 主要研究方向为网络与信息安全。



张永铮, 男, 博士, 黑龙江哈尔滨人, 中资网络信息安全科技有限公司研究员。



王佰玲 (1978-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学(威海)计算机科学与技术学院教授、博士生导师, 主要研究方向为网络安全、信息内容安全、信息对抗、工业互联网安全。



刘红日 (1982-), 男, 山东烟台人, 博士, 哈尔滨工业大学助理研究员, 主要研究方向为工业控制系统流量审计、工控互联网安全。