计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210200078

基于 DTMC 的工业串行协议状态检测算法

刘凯祥[1]　谢永芳[1]　陈　新[2]　吕　飞[2]　刘俊矫[2]
1 中南大学自动化学院　长沙 410083
2 中国科学院信息工程研究所　北京 100093
(kaixiangliu@csu.edu.cn)

摘　要　针对现有工业信息安全研究主要集中在工业以太网方面,缺少对串行链路协议防护的研究等问题,提出一种基于离散时间马尔可夫链(Discrete Time Markov Chain,DTMC)的工业串行协议状态检测算法。该算法利用工业控制系统(Industrial Control System,ICS)行为有限和状态有限的特征,根据串行链路协议历史流量数据,自动构建 ICS 正常行为模型——DTMC。模型包含状态事件、状态转移、状态转移概率和状态转移时间间隔等行为信息,使用该模型所包含的状态信息作为状态检测规则集。当检测阶段生成的状态信息与状态检测规则集中的信息不同或偏差超过阈值时,产生告警或拒绝等动作。同时,结合综合包检测(Comprehensive Packet Inspection,CPI)技术来扩大协议载荷数据的可检测范围。实验结果表明,所提算法能有效检测语义攻击,保护串行链路安全,且算法误报率为 5.3%,漏报率为 0.6%。

关键词:工业信息安全;串行链路协议;离散时间马尔可夫链;状态检测;工业控制系统;综合包检测
中图法分类号　TP393.08

Industrial Serial Protocol State Detection Algorithm Based on DTMC

LIU Kai-xiang[1],XIE Yong-fang[1],CHEN Xin[2],LYU Fei[2] and LIU Jun-jiao[2]
1 School of Automation,Central South University,Changsha 410083,China
2 Institute of Information Engineering,Chinese Academy of Sciences,Beijing 100093,China

Abstract　Aiming at the problem that the existing research on industrial security mainly focuses on industrial ethernet and lacks the research on serial link protocol protection,an industrial serial protocol state detection algorithm based on discrete time Markov chain (DTMC) is proposed. This method utilizes the characteristics of limited behavior and state of the industrial control system (ICS),and automatically constructs the normal behavior model of ICS——DTMC,based on the historical traffic data of the serial link protocol. The model contains behavior information such as state event,state transition,state transition probability and state transition time interval. Then the behavior information contained in the model is used as the state detection rule set. When the state information generated in the detection phase is different from the state detection rule set information or the deviation exceeds the threshold,actions such as alarm or rejection are generated. At the same time,combined with the comprehensive packet inspection (CPI) technology,the detectable range of protocol payload data is increased. Finally,the experimental results show that the proposed algorithm can effectively detect semantic attacks and protect the security of serial links,the false positive rate is 5.3% and false negative rate is 0.6%.

Keywords　Industrial security,Serial link protocol,DTMC,State detection,ICS,CPI

1　引言

工业控制系统(ICS)包括监控和数据采集(Supervisory Control and Data Acquisition,SCADA)系统、分布式控制系统(Distributed Control System,DCS)和可编程逻辑控制器(Programmable Logic Controller,PLC)等,广泛应用于电力、水利、石油化工和冶金等工业领域,是国家关键基础设施的重要组成部分[1-4]。近年来,为了提高远程控制的效率,ICS 与信息和通信技术广泛进行融合,但该技术为 ICS 带来高效便捷的同时也导致 ICS 易遭受网络攻击[5]。2010 年,“震网”病毒攻击伊朗核电站,损毁了大量离心机[6]。此后,“毒区”“火焰”等针对 ICS 的病毒相继爆发,使得 ICS 信息安全越来越受到重视[7]。

当前,学术界对工业防火墙与入侵检测等工业信息安全

到稿日期:2021-02-07　返修日期:2021-05-26　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金青年科学基金(61702506);国家杰出青年科学基金(61725306)
This work was supported by the Young Scientists Fund of National Natural Science Foundation of China(61702506) and National Science Fund for Distinguished Young Scholars of China(61725306).
通信作者:陈新(chenxin1990@iie.ac.cn)

防护措施展开了积极研究，但主要集中在工业以太网的安全防护上。在工业防火墙的研究上，Shang 等[8]提出使用改进的神经网络算法自学习生成白名单规则，并且准确率达到人工配置水平，但其规则仅由五元组信息组成，未对工控协议载荷内容进行检测。Pan 等[9]提取网络流量五元组信息，并对工控协议深度解析后获取功能码、合法地址等特征，使用 SVM 算法进行规则学习，也取得了较好的检测效果。Dheeraj 等[10]利用多种机器学习算法进行恶意包识别，从而动态更新防火墙规则。Yan 等[11]提出白名单分级检测的方法，综合考虑了数据流量的 ip 地址、操作码、访问地址和时间等，取得了较好的效果。但是该方法仅能防护基于时序的语义攻击，无法检测基于次序的语义攻击。入侵检测技术能通过分析采集的网络信息、检查攻击迹象等来实时保护网络，在工控系统中也得到了广泛应用[12]。Song 等[13]提出基于行为模型的异常检测算法，从 ICS 网络协议中提取行为数据序列，构建正常行为模型以预测行为数据，再将其与待检测行为数据对比分析，进行异常检测。Chen 等[14]则提出基于离群点挖掘的异常检测算法，但其仅对变量数据进行分析，未考虑操作数据，具有一定的局限性。Fovino 等[15]与 Carcano 等[16]则提出基于临界状态的入侵检测方法。以上方法都集中在工业以太网防护方面，而大部分 ICS 设备同时支持网络协议与串行链路协议，因此也应考虑工业串行链路的安全防护。

针对串行链路攻击的检测，Morris 等[17-18]使用外部计算机，将采集的 Modbus RTU 协议数据转换为 Modbus TCP 协议数据，再利用 snort 检测攻击，并且给出了针对 DoS、命令注入等攻击的检测规则，但此方案较为繁琐，需要额外的计算机进行协议转换。Tylman[19]则利用 snort 数据采集模块处理串行链路协议，将 Modbus RTU 协议数据映射到 IP 数据报文与 UDP 数据报文中，再使用 snort 进行检测，该方案不需要额外安装虚拟机进行协议转换。但该方案仅将功能码与地址码数据映射到数据报文中，而未对 Modbus RTU 协议中的其他数据进行处理。Zhang[20]则提出将 Modbus RTU 协议全部映射到 UDP 报文中，使用 snort 进行检测。

上述研究虽然对工业串行链路安全防护做出了一定贡献，但未充分考虑针对工业串行链路协议语义攻击的检测。对协议载荷的检测使用深度包检测(Deep Packet Inspection，DPI)算法，仅能检测部分载荷内容，经过特别构造、携带恶意载荷的数据包可利用该缺点绕过检测。对于 ICS 中具体的通信设备而言，往往重复固定执行有限的程序，具有通信流规律、行为特征固定和行为模式可预测，即“状态有限”和“行为有限”的特点[21]。随机过程[22]可以分析和预测系统在时空的状态变化。因此，本文结合随机过程提出一种基于离散时间马尔可夫链(DTMC)模型的工业串行链路协议状态检测算法，充分利用 ICS 行为有限与状态有限的特征，根据串行链路历史流量数据，自动构建 ICS 正常行为模型。该模型包含状态事件、状态转移、状态转移概率和状态转移时间间隔等行为信息，使用这些行为信息作为状态检测规则集。当检测阶段生成的行为信息与该状态检测规则集中的信息不同或偏差超过阈值时，产生告警或拒绝等动作。状态事件的检测结合综合包检测(CPI)算法，以扩大可检测范围。最后通过实验分析，验证了所提算法能有效检测语义攻击，保护串行链路协议信息安全。

2 相关工作

2.1 ICS 系统与串行链路协议

典型的 ICS 系统由人机界面(Human Machine Interface，HMI)、工程师站(Engineer Station，ES)、控制器(Controller)、执行器(Actuator)、传感器(Sensor)等组成，使用工业协议进行通信[14]，其模型如图 1 所示。Modbus 协议是我国标准工业通信协议，本文主要研究基于 Modbus RTU 通信的工控系统[23-24]。

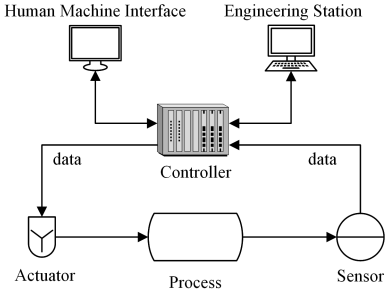

图 1 典型工控系统模型

Fig. 1 Typical industrial control system model

Modbus 协议是 Modicon 公司于 1979 年发表，位于 OSI 模型第七层的应用层协议[25]。其中，Modbus RTU 协议为串行链路协议，物理层使用 RS232 总线或者 RS485 总线；Modbus TCP 协议为以太网协议，物理层使用工业以太网。

图 2 为两种 Modbus 协议报文帧格式的对比。

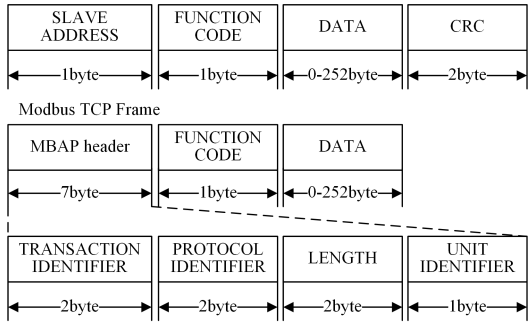

图 2 Modbus 协议帧格式对比

Fig. 2 Comparison of frame formats of Modbus protocol

Modbus RTU 协议帧最大长度为 256 字节，包括 1 字节从站地址码、1 字节功能码、0～252 字节数据及 2 字节 CRC 校验值。而 Modbus TCP 协议帧最大长度为 260 字节，包含 7 字节 Modbus 应用协议 (Modbus Application Protocol，MBAP)报文头、1 字节功能码、0～252 字节数据。除上述帧格式存在区别之外，Modbus RTU 协议与 Modbus TCP 协议的功能码也存在差别，部分关键功能码为 Modbus RTU 协议独有，仅串行链路协议使用(如功能码 8：诊断功能；功能码 17：读取从站信息)。Modbus RTU 协议地址码标识从站地址，当地址码为 0 时，该消息还表示为广播信息，所有从站必须执行且无须响应。由于这一特性，且 Modbus RTU 协议在设计之初未考虑安全性，缺少广播抑制机制，攻击者可发送

广播信息实现拒绝服务攻击[20]。此外，Modbus RTU 协议还缺乏认证机制、加密机制、权限划分等，面临其他安全风险。

2.2 CPI 算法

CPI 算法最早由 Li 等[26]提出，在传统 DPI 算法的基础上增加了对整个数据域的检测。以 Modbus RTU 协议为例，DPI 算法与 CPI 算法可检测区域的对比如图 3 所示。

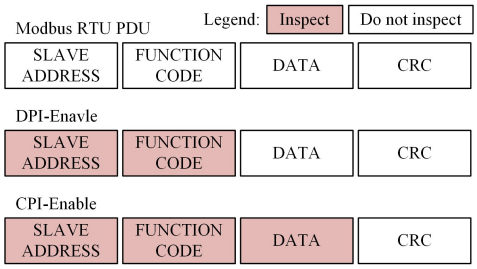

图 3 DPI 算法与 CPI 算法检测区域的对比

Fig. 3 Comparison of inspection areas between CPI and DPI

为解决不同协议的载荷结构存在差异性等难题，CPI 算法将载荷内容按层次划分，实现更加系统、全面、有效的检测。图 4 以一条具体的 Modbus RTU 消息为例，说明各个域的位置。

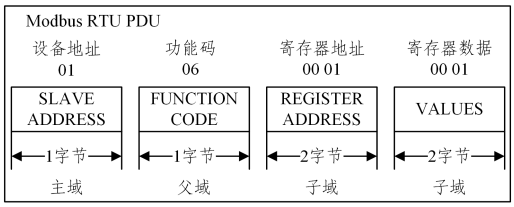

图 4 CPI 各域位置

Fig. 4 Location of each CPI field

协议载荷内容的层次如下：1）主域，载荷中首先检测的域，如 Modbus RTU 协议中的从站地址；2）父域与子域，载荷中有关联性的相邻域，如 Modbus RTU 协议中的功能码域与数据域，前者为父域，后者为子域；3）独立域，该域与其他域相互无关联，如标志数据；4）固有范围与限定范围，如功能码或寄存器地址等在协议标准中已规定范围，即为固有范围。但在实际设备中，不同产品支持的功能码或寄存器地址范围不同，其支持的功能码或寄存器地址范围为限定范围。

2.3 攻击模型

ICS 所面临的主要攻击有注入攻击与语义攻击[15,17-18,27-28]。注入攻击[17-18]包括响应注入与命令注入。响应注入为攻击者在 ICS 中注入虚假响应数据包。ICS 反馈控制决策依赖于控制系统过程数据，攻击者可以通过注入虚假的响应数据包使得控制算法或操作员做出错误的控制决策。命令注入为攻击者在 ICS 中注入非法控制命令或配置命令。命令注入攻击有两种方式，一种是攻击者将错误的控制操作注入 ICS 中，另一种是攻击者覆盖远程终端等可编程设备中梯形图或寄存器所包含的关键控制参数，执行错误操作。语义攻击是一类特殊的语义学攻击，包括次序攻击与时序攻击。语义攻击发送的数据包序列，从单条数据包来看为合法数据包，但这些数据包按特殊的次序或时序发送，迫使系统处于危险状态[27]。次序攻击是攻击者将正常的消息或命令序列以错误的顺序发送，如 Fovino 等[15]选用一条输气管道，通过两条错位的合法控制命令，完全打开输气开关并完全关闭进气开关，从而使得输气量及管道内压强最大化，导致系统进入危险状态。时序攻击是将正常的消息或命令序列在错误的时间发送，如美国总统关键基础设施保护委员会的报告中提到[28]，在输水管道中，通过频繁开合控制阀门，引发水锤效应，导致大量管道同时破裂。

3 方法介绍

3.1 DTMC

马尔可夫模型是经典的概率统计模型[29]，本文应用 DTMC 模型进行状态检测。下面给出 DTMC 的相关定义与引理[30]。

定义 1 如果一个随机过程$\{X_n, n\geqslant 0\}$为具有状态空间$\boldsymbol{S}=\{s_0, s_1, \cdots\}$的(一阶)DTMC。则对所有的 $n\geqslant 1$，以及任意状态 $j\in\boldsymbol{S}$ 与 $s_m\in\boldsymbol{S}(0\leqslant m\leqslant n)$，式(1)成立。

$$\Pr(\boldsymbol{X}_{n+1}=j\mid\boldsymbol{X}_n=s_n,\boldsymbol{X}_{n-1}=s_{n-1},\cdots,\boldsymbol{X}_0=s_0)=\Pr(\boldsymbol{X}_{n+1}=j\mid\boldsymbol{X}_n=s_n) \tag{1}$$

定义 2 一阶 DTMC$\{\boldsymbol{X}_n, n\geqslant 1\}$对于所有 $j\in\boldsymbol{S}$ 和 $i\in\boldsymbol{S}$，若条件概率 $\Pr(\boldsymbol{X}_{n+1}=j\mid\boldsymbol{X}_{n=i})$与 n 无关，则其满足不动性假设，称作是时间齐次的。

当 DTMC$\{X_n, n\geqslant 1\}$是时间齐次的，$p_{ij}=\Pr(\boldsymbol{X}_{n+1}=j\mid\boldsymbol{X}_n=i)$称作(一步)转移概率，由(一步)转移概率 p_{ij} 构成的矩阵 $\boldsymbol{P}=[p_{ij}]$称作(一步)转移概率矩阵。$a_i=\Pr(\boldsymbol{X}_0=i)$称作状态 i 的初始出现概率，由状态的初始出现概率构成的行向量 $\boldsymbol{A}=(a_i)_{i\in S}$称作初始概率分布。

引理 1[30] 时间齐次的 DTMC$\{\boldsymbol{X}_n, n\geqslant 0\}$由初始概率分布 $\boldsymbol{A}$ 和转移概率矩阵 $\boldsymbol{P}$ 完全刻画，即：

$$\Pr(\boldsymbol{X}_0=s_0,\cdots,\boldsymbol{X}_{n-1}=s_{n-1},\boldsymbol{X}_n=s_n)=a_{s_1}p_{s_1,s_2},\cdots,p_{s_{n-1},s_n} \tag{2}$$

定义 3 如果一个 DTMC$\{\boldsymbol{X}_n, n\geqslant 0\}$初始概率分布 $\boldsymbol{A}$ 和转移概率矩阵 $\boldsymbol{P}$ 满足式(3)，则称之为平稳的。

$$\boldsymbol{A}=\boldsymbol{A}\times\boldsymbol{P} \tag{3}$$

引理 2[30] 对于平稳的 DTMC$\{\boldsymbol{X}_n, n\geqslant 0\}$，式(4)成立：

$$\Pr(\boldsymbol{X}_n=i)=\Pr(\boldsymbol{X}_1=i)=a_i \tag{4}$$

由引理 1 和引理 2 得引理 3。

引理 3[30] 对于平稳的、时间齐次的 DTMC$\{\boldsymbol{X}_n, n\geqslant 0\}$，式(5)成立：

$$\Pr(\boldsymbol{X}_i=s_i,\cdots,\boldsymbol{X}_{j-1}=s_{j-1},\boldsymbol{X}_j=s_j)=a_{s_i}p_{s_i,s_{i+1}},\cdots,p_{s_{j-1},s_j},0\leqslant i<j \tag{5}$$

3.2 模型训练

3.2.1 状态事件定义

本文方法使用 DTMC 模型对 ICS 正常行为进行建模，DTMC 的状态空间为 $\boldsymbol{S}=\{1,2,\cdots,N\}$，根据指令信息确定具体状态。以 Modbus RTU 协议为例，对每一条具体的指令消息，可获取六元组信息〈slaveId, pktType, funCode, pduData, checkSum, time〉，使用该六元组定义状态事件。其中：

1)slaveId 表示设备从站地址;2)pktType 表示报文请求/响应类别;3)funCode 表示协议功能码;4)pduData 表示协议参数数据,如线圈/寄存器地址、线圈/寄存器数量等;5)checkSum 表示校验码数据;6)time 表示状态事件时间戳。

通过该六元组信息,Modbus RTU 指令序列映射的状态事件皆唯一。此外,每个状态事件还包含两个统计属性:状态事件出度值 OUT 与状态事件入度值 IN。

3.2.2 构建 DTMC 模型

使用捕获的串行链路协议历史数据构建 DTMC 模型,具体步骤如下。

Step1 根据状态事件的六元组定义完成指令数据到状态事件的映射,若映射成功,则直接执行 Step3,否则执行 Step2。

Step2 根据当前消息创建新的状态事件,执行 Step3。

Step3 更新当前状态与前一跳状态的转移关系,将前一跳状态事件出度值加 1,当前状态事件入度值加 1;将当前状态事件 time 值设为当前消息的时间戳,执行 Step4。

Step4 将当前消息时间戳与前一跳状态事件时间戳的差值作为转移时间间隔信息添加至对应的转移关系序列集合中,重复执行 Step1 直到所有历史数据映射完毕,执行 Step5。

Step5 根据状态转移关系信息及状态出度值信息等计算状态转移概率与转移时间间隔均值。

为了详细说明 DTMC 的构建过程,以如下 4 条通信数据为例进行说明。

(1)slaveId="1",pktType="0",funCode="1",pduData="0,2",checkSum="52157",time="2021-1-11 10:14:00.269"。

(2)slaveId="1",pktType="1",funCode="1",pduData="0,2;00",checkSum="34897",time="2021-1-11 10:14:00.347"。

(3)slaveId="1",pktType="0",funCode="1",pduData="0,2",checkSum="52157",time="2021-1-11 10:14:01.268"。

(4)slaveId="1",pktType="1",funCode="1",pduData="0,2;00",checkSum="34897",time="2021-1-11 10:14:01.346"。

利用上述通信数据构建 DTMC 模型的第一阶段:对于通信数据(1),数据到状态事件的映射失败,因为当前未创建任何状态事件。因此创建新的状态事件 1,该状态事件表示一条主站发送至从站地址为 1 的设备,读取线圈地址 0 开始的 2 个值的请求消息。通信数据(2),表示一条从站地址 1 的设备,返回线圈地址 0 开始的 2 个值的响应消息。该消息无法映射到状态 1,因此创建新的状态 2,将该消息时间戳"2021-1-11 10:14:00.347"与状态事件 1 时间戳"2021-1-11 10:14:00.269"的差值添加到对应边的数值集合中,并将对应边的 count 值加 1,将状态事件 1 的出度值加 1,将状态事件 2 的入度值加 1。具体过程如图 5 所示。

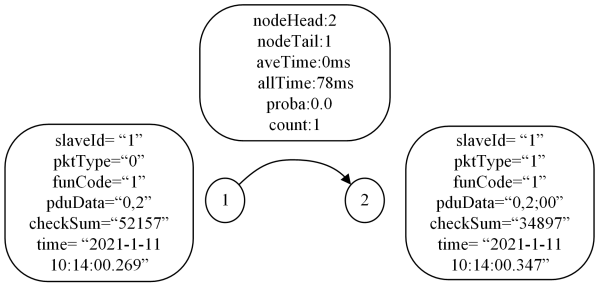

图 5 DTMC 模型构建阶段 1

Fig. 5 DTMC model building phase 1

利用上述通信数据构建 DTMC 模型的第二阶段:通信数据(3)根据状态事件的定义可映射到状态事件 1,因此将状态事件 1 的 time 值设为当前消息的时间戳"2021-1-11 10:14:01.268",将该数据时间戳与状态事件 2 时间戳的差值添加到对应边的数值集合中,并将对应边的 count 值加 1,将状态事件 2 的出度值加 1,将状态事件 1 的入度值加 1;通信数据(4)根据状态事件的定义可映射为状态事件 2,因此将状态事件 2 的 time 值设为当前消息的时间戳"2021-1-11 10:14:01.346",将该数据时间戳与状态事件 1 时间戳的差值添加到对应边的数值集合中,并将对应边的 count 值加 1,将状态事件 1 的出度值加 1,将状态事件 2 的入度值加 1。具体过程如图 6 所示。

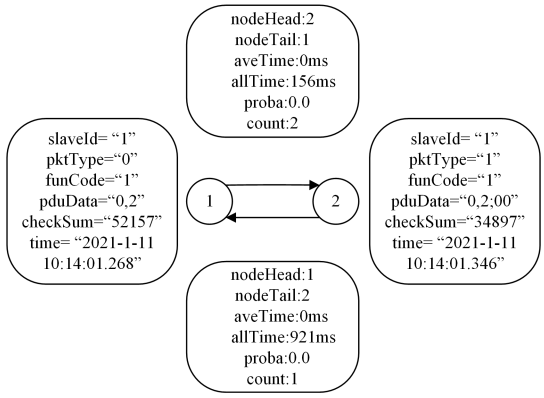

图 6 DTMC 模型构建阶段 2

Fig. 6 DTMC model building phase 2

所有历史信息数据映射完毕后,需对每个状态节点的转移关系进行统计,计算转移概率以及转移时间间隔均值。以〈1,2〉转移关系为例,记状态节点 1 的出度为 N,边序列转移次数为 n,时间间隔累计和为 τ,则状态节点 1,2 间的转移关系中转移概率 ρ 的计算式如式(6)所示,时间间隔均值 δ 的计算式如式(7)所示。

$$\rho=n/N \tag{6}$$

$$\delta=\tau/n \tag{7}$$

假设利用 3.2.2 节中的 4 条数据构建 DTMC,〈1,2〉转移关系中,状态节点 1 出度为 2,〈1,2〉边序列转移次数为 2,时间间隔累计和为 156 ms,则其对应的转移概率为 1.0,时间间隔均值为 78 ms。

3.3 检测过程

使用构建完成的 DTMC 模型进行检测,可根据状态信息检测出非法状态节点、非法状态转移、非法状态转移概率以及

非法状态转移时间间隔。状态转移概率需要一定数量的历史数据才能得到稳定值，据此，将检测过程分为两个阶段，即起始阶段与后续阶段。起始阶段仅检测状态节点、状态转移和状态转移时间间隔的合法性。获取一定数量历史数据后，增加对状态转移概率的检测，可根据 ICS 业务流程的周期性设定该数量。检测流程图如图 7 所示。

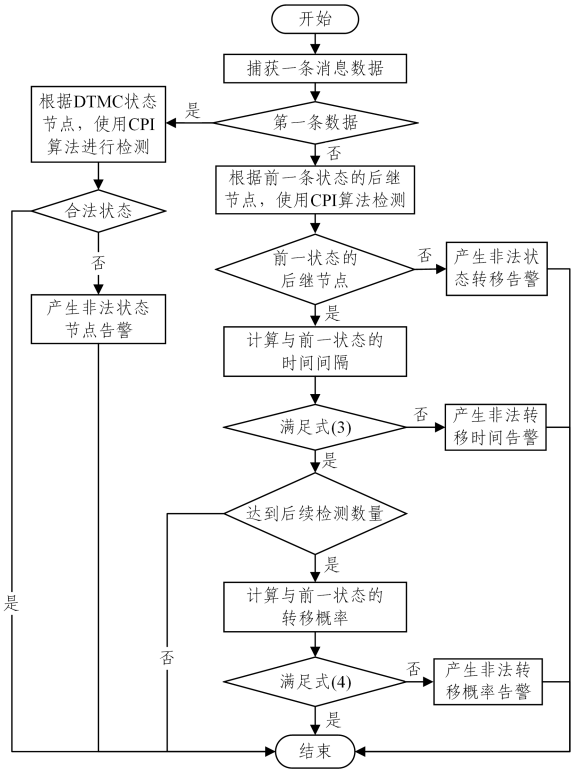

图 7 检测流程图

Fig. 7 Detection process

具体检测步骤如下所述。

(1)起始阶段

Step1 对第一条消息数据，根据已构建的 DTMC 模型节点，使用 CPI 算法检测消息数据，若节点中未检测到该信息数据，则产生“非法状态节点”告警；否则信息数据为合法数据，对后续数据执行 Step2。

Setp2 根据前一状态节点的后继节点信息，使用 CPI 算法检测当前消息数据，若未检测到该信息数据，则产生“非法状态转移”告警；若检测到则执行 Step3。

Step3 计算当前状态与前一状态的转移时间间隔 τ_{detect}，执行 Step4。

Step4 将 τ_{detect} 与对应的 τ_{train} 使用式(8)进行判断，若式(8)成立，则产生“非法转移时间间隔”告警，其中 θ_τ 为时间间隔偏差阈值。

$$\frac{|\tau_{\text{detect}}-\tau_{\text{train}}|}{\tau_{\text{train}}}>\theta_\tau \tag{8}$$

(2)后续阶段

Step1 计算当前状态与前一状态的关系转移概率 ρ_{detect}，执行 Step2。

Step2 将 ρ_{detect} 与对应的 ρ_{train} 使用式(9)进行判断，若式(9)成立，则产生“非法转移概率”告警，其中 θ_ρ 为概率偏差阈值。

$$|\rho_{\text{detect}}-\rho_{\text{train}}|>\theta_\rho \tag{9}$$

4 仿真实验与结果分析

4.1 模拟仿真实验

4.1.1 实验环境

为验证本文所提模型的有效性，首先进行模拟仿真实验，利用 Modbus Poll/Slave 仿真软件搭建一个小型仿真系统，系统结构如图 8 所示。

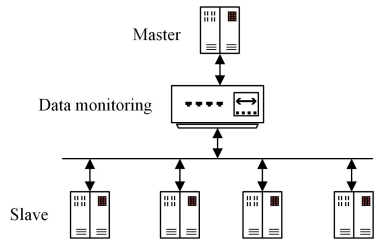

图 8 软件仿真系统架构图

Fig. 8 Software simulation system architecture

主站使用 Modbus Poll 软件仿真，设置 3 种指令发送时间：1 s，2 s 和 10 s。使用 Modbus Slave 软件模拟 12 个从站设备。

4.1.2 实验数据生成

实验阶段，采集该系统连续运行 14 h 产生的流量数据，保证充分获取整个系统的行为信息。选取前 12 h 的数据作为正常行为数据，后 2 h 的数据通过篡改、重放等[17-18,27-28]方式生成攻击数据，作为模型检测数据。其中，注入攻击数据集为 attack1，该攻击篡改线圈/寄存器地址或数量；周期性时序攻击数据集为 attack2，该攻击通过重放、缩短数据请求/应答时间实现；次序攻击数据集为 attack3，该攻击篡改、重放数据包的次序。具体待检测数据类别如表 1 所列。

表 1 检测数据类型

Tabel 1 Test data types

数据集	数据条数
normal	6 064
attack1	2 000
attack2	1 000
attack3	965

4.1.3 实验结果与分析

将本文算法与 STG 算法[7]、决策树(Decision Tree，DT)算法[10]和卷积神经网络(Convolutional Neural Networks，CNN)算法[31]进行对比。使用正常数据集训练 STG 算法[7]，攻击数据集用于检测。对于 DT 算法[10]与 CNN 算法[31]，增加正常数据集与攻击数据集标签，使用滑动窗口[32]对数据集进行处理，窗口内数据全为正常数据，增加正常数据标签，否则增加攻击数据标签。检测结果如表 2 所列。

表 2 算法检测结果对比

Tabel 2 Comparison of detection results of four algorithms

Algorithm	attack1	attack2	attack3
DTMC	检测成功	检测成功	检测成功
STG	检测成功	检测失败	检测成功
CNN	检测成功	检测成功	检测成功
DT	检测成功	检测成功	检测成功

通过表 2 可知，DTMC 算法、带滑动窗口的 CNN 算法、

DT 算法与 STG 算法相比都具有更完整的检测能力，除能检测注入攻击与次序攻击外，还能检测时序攻击。DTMC 算法与带滑动窗口的 CNN 算法、DT 算法相比，使用正常数据构建模型，即可对攻击数据进行有效检测，而 CNN 算法与 DT 算法需要额外使用攻击数据进行训练，才能对攻击数据进行有效检测。此外，ICS 流量数据中，攻击数据相对难以获取，仅使用正常流量数据构建模型的算法更适合工业信息安全防护的应用场景。

4.2 实物仿真实验

4.2.1 实验环境

为进一步验证本文算法的有效性，使用真实系统进行实验。本实验搭建一个小型控制系统，该系统属于交通信号灯控制系统。该系统中，人机界面使用昆仑通态 TPC1061Ti，控制器为和利时 PLC，LM3109 CPU，通信协议为 Modbus RTU 协议。人机界面为主站，和利时 PLC 为从站，通过人机界面读取和控制和利时 PLC，信号灯与 PLC 输出点位连接，模拟红绿灯与计时器。系统架构如图 9 所示。

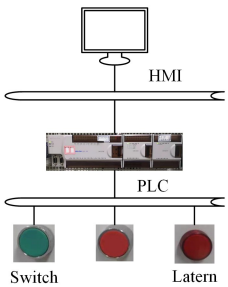

图 9 交通信号灯仿真系统

Fig. 9 Traffic signal simulation system

4.2.2 实验数据生成

本次实验中，计时器循环时间为 31 s，为保证模型能够充分学习到整个系统的串口流量特征，采集系统稳定运行长达 1 h 产生的流量数据。其中，使用前 50 min 的流量数据作为学习数据，构建 DTMC 模型，后 10 min 的流量数据通过篡改、重放等方式生成攻击数据集，数据类别如表 3 所列。

表 3 攻击数据

Tabel 3 Attack data

攻击类型	数据
Normal	5418
Attack1	2000
Attack2	1000
Attack3	115

数据集中，注入攻击数据集为 attack1，周期性时序攻击数据集为 attack2，次序攻击数据集为 attack3。

4.2.3 实验结果与分析

为更加全面地评估检测效果，使用误报率与漏报率两种指标进行分析[3]。误报率是将正常行为误判断为异常行为的概率，漏报率是将异常行为误判断为正常行为的概率，其计算公式如下：

$$FPR=FP/(FP+TP) \tag{10}$$

$$FNR=FN/(FN+TN) \tag{11}$$

其中，FPR 为误报率，FNR 为漏报率，TP 为正常数据被允许通过的数量，FP 为正常数据产生告警的数量，TN 为攻击数据产生告警的数量，FN 为攻击数据被允许通过的数量。

该实验中，根据先验知识以及经验，DTMC 模型检测周期为 280 条数据，概率异常阈值为 0.1，时间间隔异常阈值为 0.2。CNN 算法选用 1 层卷积层、1 层池化层，隐藏单元数为 8。使用滑窗处理数据时窗口大小为 280，步长为 1。最后，各个算法的检测结果如表 4、表 5 所列。

表 4 算法检测结果对比

Table 4 Comparison of detection results of four algorithms

(单位：%)

Algorithm	*FPR*	*FNR*
DTMC	5.3	0.6
STG	0.1	35.1
CNN	9.9	3.2
DT	7.6	1.0

表 5 攻击检测结果对比

Table 5 Comparison of attack detection results of four algorithms

(单位：%)

Algorithm	attack1	attack2	attack3
DTMC	100	99.1	99.1
STG	100	0.6	13.9
CNN	100	100	11.0
DT	100	100	73.9

通过表 4 可知，本文算法与 CNN，DT 算法相比，误报率与漏报率都较低。而通过表 5 可知，本文算法与 CNN，DT 算法相比，对 attack3 攻击的检测效果具有明显优势。在表 4 中，本文算法与 STG 算法相比，误报率略高，但漏报率远低于 STG 算法。由表 5 可以看出，STG 算法难以检测时序攻击与次序攻击。通过对告警结果分析发现，STG 算法产生的时序攻击告警信息与次序攻击告警信息是模型构建阶段未学习到的转移关系。这主要因为本文算法相较于 STG 算法增加了对转移时间间隔与转移概率信息的检测。综上，本文算法的误报率和漏报率都较低，与其他 3 种算法相比更具有优势。

结束语 本文针对现有工业信息安全研究主要集中在工业以太网方面，缺少对串行链路协议防护的研究等问题，提出一种基于离散时间马尔可夫链(DTMC)的工业串行协议状态检测算法。该算法充分利用 ICS 行为有限和状态有限的特征，根据串行链路协议历史流量数据，自动构建 DTMC 模型，通过引入时间间隔信息，增强指令间的关系，来检测时序攻击。同时，结合 CPI 算法增加工控协议载荷可检测内容。最后，通过实验分析，验证了所提算法能有效检测语义攻击，保护串行链路协议信息安全。但该方法也存在一些不足，如未考虑非周期数据，无法同时检测串行链路协议数据与网络协议数据。在未来的研究中，我们将进一步考虑增加对非周期数据的检测，以及对串行链路协议与网络协议整体的保护。

参 考 文 献

[1] LAI Y, LIU Z, LIU J. Abnormal detection method of industrial control system based on behavior model[J]. Computers & Security, 2019, 84(JUL.): 166-178.

[2] SUO Y F, WANG S J, QIN Y, et al. Summary of Security Technology and Application in Industrial Control System[J]. Computer Science, 2018, 45(4): 25-33.

[3] YANG A，SUN L M，WANG X S，et al. Intrusion detection techniques for industrial control systems[J]. Journal of Computer Research and Development，2016，53(9)：2039-2054.

[4] GUO X，WANG Y Y，FENG T，et al. Blockchain-based Role-Delegation Access Control for Industrial Control System[J]. Computer Science，2021，48(9)：306-316.

[5] FENG C，LI T，CHAN A D. Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks[C]// 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. IEEE，2017：261-272.

[6] LANGNER R. Stuxnet：dissecting a cyberwarfare weapon[J]. IEEE Security and Privacy，2011，9(3)：49-51.

[7] LV X F，XIE Y B. An Anomaly Detection Method for Industrial Control Systems via State Transition Graph[J]. Acta Automatica Sinica，2018，44(9)：1662-1671.

[8] LEI Y Q，SHANG W L，WAN M，et al. Industrial firewall rules self-learning algorithm design[J]. Computer Engineering and Design. 2016，37(12)：613-617.

[9] PAN F，WANG S W，XUE P. Self-learning method of industrial firewall rules based on SVM algorithm[J]. Information Technology and Network Security，2018，37(5)：29-33.

[10] DHEERAJ R，GUO H，VEERAVALLI B，et al. Design and Development of SCADA Firewall Security Features for Protecting Industrial Operations[C]// 2019 IEEE VTS Asia Pacific Wireless Communications Symposium. IEEE，2019：1-5.

[11] YAN B，YIN L B，YING H，et al. Hierarchical Intrusion Detection Algorithm based on White List for Industrial Control Network[J]. Communication Technology，2018，51(4)：907-912.

[12] LU Y. Research on a New Hybrid Intrusion Detection Algorithm for Cloud Computing[J]. Journal of Chongqing University of Technology (Natural Science)，2020，34(10)：153-159.

[13] SONG Z W，ZHOU R K，LAI Y X，et al. Anomaly Detection Method of ICS Based on Behavior Mode[J]. Computer Science，2018，45(1)：233-239.

[14] CHEN Z，HUANG Y，ZOU H. Anomaly Detection of Industrial Control System Based on Outlier Mining[J]. Computer Science，2014，41(5)：178-181，203.

[15] FOVINO I N，CARCANO A，MUREL T D L，et al. Modbus/DNP3 state-based intrusion detection system[C]// 2010 24th IEEE International Conference on Advanced Information Networking and Applications. IEEE，2010：729-736.

[16] CARCANO A，COLETTA A，GUGLIELMI M，et al. A Multidimensional Critical State Analysis for Detecting Intrusions in SCADA Systems[J]. IEEE Transactions on Industrial Informatics，2011，7(2)：179-186.

[17] MORRIS T，VAUGHN R，DANDASS Y. A retrofit network intrusion detection system for MODBUS RTU and ASCII industrial control systems[C]// 2012 45th Hawaii International Conference on System Sciences. IEEE，2012：2338-2345.

[18] MORRIS T，JONES B，VAUGHN R，et al. Deterministic intrusion detection rules for MODBUS protocols[C]// 2013 46th Hawaii International Conference on System Sciences. IEEE，2013：1773-1781.

[19] TYLMAN W. Native support for Modbus RTU protocol in Snort intrusion detection system[M]// New Results in Dependability and Computer Systems. Heidelberg：Springer，2013：479-487.

[20] ZHANG Y. Research on Industrial Control System Intrusion Detection Technology[D]. Chengdu：University of Electronic Science and Technology of China，2018.

[21] SHANG W L，QIAO Q S，WAN M，et al. Self-learning method for generation and optimization of industrial firewall rules[J]. Computer Engineering and Design，2016，37(7)：1752-1756.

[22] ROSS S M. Introduction to Probability Models[M]. Ninth Edition. Singapore：Elesevier，2007：185-263.

[23] GB/T 19582.1-2008. Modbus industrial automation network specification—Part 1：Modbus application protocol[S]. Beijing：China Standard Press，2008.

[24] GB/T 19582.2-2008. Modbus industrial automation network specification-Part 2：Modbus protocol implementation guide over serial link[S]. Beijing：MarkovChain，2008.

[25] MODBUS IDA. MODBUS over Serial Line Specification and Implementation Guide v1. 02[EB/OL]. http://www.modbus.org/docs/Modbus_over_serial_line_V1_02.pdf，December 20，2006.

[26] LI D，GUO H，ZHOU J，et al. SCADAWall：A CPI-enabled firewall model for SCADA security[J]. Computers & Security，2019，80(JAN.)：134-154.

[27] CASELLI M，ZAMBON E，KARGL F. Sequence-aware intrusion detection in industrial control systems[C]// Proceedings of the 1st ACM Workshop on Cyber-Physical System Security. 2015：13-24.

[28] FOUNDATIONS C. Protecting America's Infrastructures：The Report of the President's Commission on Critical Infrastructure Protection[R]. Washington DC：The President's Commission on Critical Infrastructure Protection.

[29] ZHANG J T，ZHOU J，XU H L，et al. An Arterial Travel Time Estimation Model Based on Discrete Time Markov Chains[J]. System Engineering，2014，32(5)：98-104.

[30] KARLIN S，TAYLOR H. A First Course in Stochastic Processes[M]. Second Edition. Beijing：Posts & Telecom Press，2007.

[31] ZHAO Z Y，XIA X J. Intrusion Detection Algorithm of Power Grid Industrial Control System Based on CNN[J]. Computer Systems & Applications，2020，29(8)：179-184.

[32] SHANG W L，ZHANG S S，WAN M，et al. Modbus/TCP Communication Anomaly Detection Algorithm Based on PSO-SVM[J]. Acta Electronica Sinica，2014，42(11)：2314-2320.

LIU Kai-xiang，born in 1995，postgraduate. His main research interests include security of industrial control systems and so on.

CHEN Xin，born in 1990，master，intermediate engineer. His main research interests include ICS security and ICS intrusion detection.

（责任编辑：李亚辉）