基于 PSO-SVM 的 Modbus TCP 通讯的异常检测方法

尚文利[1,3],张盛山[1,2,3],万 明[1,3],曾 鹏[1,3]
(1.中国科学院沈阳自动化研究所,辽宁沈阳 110016; 2.中国科学院大学,北京 100039;
3.中科院网络化控制系统重点实验室,辽宁沈阳 110016)

摘 要: 如何有效检测和防御工业病毒对应用层协议数据的攻击是目前工业安全网关研究的难点问题.本文提出了将 Modbus TCP 通讯流量转换为异常检测模型所需数据形式的预处理方法,设计了一种利用粒子群 PSO 算法进行参数寻优的 PSO-SVM 算法.该方法根据 Modbus 功能码序列中的模式短序列出现的频率,识别出异常的 Modbus TCP 通讯流量.最后,通过实验数据分析,说明了提出方法可以有效实现对 Modbus 功能码序列的异常检测.

关键词: 微粒子群; 支持向量机; Modbus 功能码; 序列异常检测; 工业安全网关
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112 (2014)11-2314-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2014.11.029

Modbus/TCP Communication Anomaly Detection Algorithm Based on PSO-SVM

SHANG Wen-li[1,3], ZHANG Sheng-shan[1,2,3], WAN Ming[1,3], ZENG Peng[1,3]
(1. *Shenyang Institute of Automation, Chinese Academy of Science, Shenyang, Liaoning* 110016, *China*;
2. *University of Chinese Academy of Sciences, Beijing* 100039, *China*;
3. *Networked Control Systems Key Laboratory of CAS, Shenyang, Liaoning* 110016, *China*)

Abstract: To detect and defend industry virus attacks to application layer protocol data is difficult issues in study of industrial security gateway. In this paper, a data pre-processing method is presented, which can convert Modbus TCP traffic into anomaly detection model, and a PSO-SVM algorithm is designed, which optimizes parameters by advanced Particle Swarm Optimization (PSO) algorithm. The method identifies anomalies of Modbus TCP traffic according to appear frequencies of the mode short sequence of Modbus function code sequence. Finally, experimental data analysis shows that the proposed method can effectively detect abnormal of Modbus function code sequence.

Key words: PSO; SVM; Modbus function code; sequence anomaly detection; industrial security gateway

1 引言

工业控制系统在设计之初由于普遍采用专有的通信协议、操作系统、硬件设备,并且与其他网络隔离,更多关注的是物理安全和功能安全,欠缺信息与网络安全方面的考虑.伴随着信息化的需求,工业控制系统的封闭性正在不断被打破,TCP/IP 技术、开放的工业通讯协议、通用操作系统等应用得越来越广泛,使得"天生"存在很多信息与网络安全缺陷的工业控制系统更加脆弱[1~4].

Modbus TCP 协议广泛应用于石油化工、能源、冶炼、电力等工业控制系统和 SCADA 系统中,国内外已有学者、机构针对其脆弱性分析、攻击与安全防护等方面进行了一定的研究工作,目前相关的代表性研究成果有:孙大林等[5]对 Modbus TCP 在工业监控系统中的安全性进行了分析,并且提出了诸多技术与管理层面上的解决方案,可以有效降低工控系统信息安全风险,但该研究主要是将传统 IT 网络中的安全策略迁移到工业控制系统中,缺乏针对 Modbus 应用层协议的研究.王婷婷[6]设计了一种改进的 DES 加密算法对 RTU 模式下的 MODBUS 通讯协议进行加密传输,但是加密/解密的方法要消耗大量的资源与时间,该研究缺乏对实时性能方面的论证.张云贵等[7]提出了基于工业控制模型的非参数累积和 CUSUM 入侵方法,该方法计算工业控制模型的预测输出与传感器测量信号的差值,形成基于时间的统计序列,采用 CUSUM 算法实现在线入侵监测及报警,但是该方法首先需要建立在熟悉工业控制系统工艺流程的基础之上,其次对建立的工业控制模型本身就可能存在

收稿日期:2013-08-30;修回日期:2014-04-12;责任编辑:蓝红杰
基金项目:国家自然科学基金(No.61164012);国家 863 高技术研究发展计划(No.2012AA041102-03)

较大误差.JavierJiménez 等[8]提出了利用 Modbus TCP 攻击行为的签名进行工业控制系统的入侵检测,基于签名方式的入侵检测系统可以有效识别入侵行为并进行报警提示,但其准确性受到数据签名库的制约,签名的不完备性决定了该方法无法识别所有攻击行为.VenkatPothamsetty 等[9]提出利用 Linux 内核中 Netfilter/Iptables 框架实现 Modbus 应用层协议字段过滤的防火墙技术,Modbus TCP 应用层防火墙的规则配置很难做到拦截一切攻击行为而不影响正常的工业通讯行为.在实际应用方面,面向工业控制系统和 SCADA 系统的工业级防火墙,采用了深度包检查技术(Deep Packet Inspection,简称 DPI),用于深入解析 Modbus TCP 应用层协议数据[10].

综上,现有的防护方法从安全策略、传输加密、协议应用层的入侵检测与访问控制等几个方面形成了 Modbus TCP 安全防护体系,但是存在一个主要的缺失,即为无法辨识出未知特征的攻击或入侵行为,也不能拦截利用防火墙规则配置发起的攻击行为.为解决该问题,本文选取 Modbus 功能码这一重要字段作为研究对象,结合支持向量机算法,提出了一种基于支持向量机的 Modbus TCP 通讯功能码序列异常检测方法,并采用微粒子群算法对模型参数进行寻优,建立了工业控制系统中 Modbus TCP 通讯的分类模型,以实现对防火墙与入侵检测系统未能识别的攻击行为或者异常行为的辨识.

2　Modbus TCP 通讯序列的特征选择

Modbus 是 OSI 模型第七层上的应用层报文传输协议,已经成为一种通用的标准.其实现方式主要包括基于串行链路的实现(Modbus RTU/Modbus ASCII)、基于高速令牌环的实现(Modbus Plus)和基于以太网上 TCP/IP 技术的实现(Modbus TCP/IP,以下简称 Modbus TCP).本文主要关注 Modbus TCP 的通讯安全,其请求与响应被封装为如图 1 所示的报文格式[11,12].

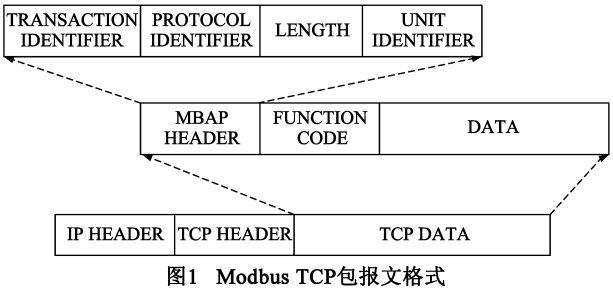

图1　Modbus TCP包报文格式

其中 MBAP HEADER 为 Modbus 应用协议报文头,目的是识别 Modbus 应用数据单元.该部分主要包含事务处理标志符(TRANSACTION IDENTIFIER)、协议标志符(PROTOCOL IDENTIFIER)、长度(LENGTH)和单元标志符(UNIT IDENTIFIER)四个部分[11].功能码是 Modbus 客户端(SLAVE)向服务端(MASTER)指示进行何种操作的标志字段,最能够体现客户端对服务器的操作意图.数据部分(DATA)由客户机根据具体的应用和功能码的不同进行设置,并由服务端进行相应的回答.

Modbus TCP 协议底层是基于标准的以太网 TCP/IP 技术[12],因此底层协议的安全缺陷被继承保留下来.除此之外,Modbus 协议应用层设计也存在缺陷,主要表现为缺乏认证、授权和加密等安全机制.缺乏认证表现在仅需要一个合法的 Modbus 地址与合法的功能码即可以建立一个 Modbus 会话;缺乏授权体现在没有基于角色的访问控制机制,任意用户可以执行任意的功能;缺乏加密体现在地址和命令使用明文传输,很容易捕获并进行解析[13,14].例如,遭受病毒感染的 Modbus 客户端可以监听并解析网络中的通讯流量,利用侦听得到的功能码和线圈或寄存器构造恶意数据包,即可修改 Modbus 服务器端中的线圈值或寄存器值,造成严重的后果,而这种攻击行为是无法通过 Modbus TCP 应用层过滤的工业防火墙进行拦截,因为防火墙规则设置无法将该攻击流量从 Modbus TCP 正常通讯的流量中分离出来.

工业控制系统与 SCADA 系统在投入运行前,需要根据生产过程与工艺要求进行编程组态,需要在不同的时间节点执行事先指定的动作,具有明显周期特性.可以推测,Modbus TCP 系统在稳定运行时的网络通讯具有一定的序列特征和行为模式.而遭受病毒感染的客户端向服务器端发起攻击,必然要产生 Modbus 报文,从而破坏 Modbus TCP 系统稳定的序列特性.因此,可对正常运行的 Modbus TCP 系统建立正常网络通讯行为模型,从而识别出异常通讯行为或者攻击行为.

使用 Modbus 协议的功能码字段,可以有效表征客户端对服务器端的操作意图[11],因此本文选择 Modbus 功能码作为研究对象,将连续的 Modbus TCP 通讯网络数据包抽象简化为 Modbus 功能码序列,从而将 Modbus TCP 系统网络通讯异常检测转化为 Modbus 功能码序列的异常检测.

3　数据预处理

工业控制系统中获得的 Modbus TCP 流量并不能直接用来进行异常检测,需要经过一系列的预处理流程.首先,在保持通讯数据包的时间先后顺序的情况下,随机将 Modbus TCP 数据包流分割成不同的序列,构成了两类样本的雏形.然后,赋予不同的数据包的类别标签,凡是包含模拟攻击源的 Modbus TCP 数据包的通讯序列即判定为“异常”,凡是不包含模拟攻击源的数据

包的通讯序列判定为“异常”.最后,去除两类样本中的除 Modbus 功能码之外的其他非必要信息,即可得到 Modbus 功能码序列样本集合.

然而,对于 Modbus 功能码序列的异常检测却不能直接进行,因为支持向量机 SVM 等智能算法处理的样本数据需要具有相同维数,而经过随机截断的 Modbus 功能码序列可能包含不同数目的 Modbus 功能码.显然,将单个功能码简单地映射为样本某一维度数值的方式并不可行.当然,如果每一个 Modbus 功能码序列都采用具有相同长度的功能码可以避免该问题,但却丧失了样本选择的随机性和灵活性.

本文提出一种将包含不同数目 Modbus 功能码序列转化为相同长度向量的方法,算法流程如图 2 所示.其具体过程如下:

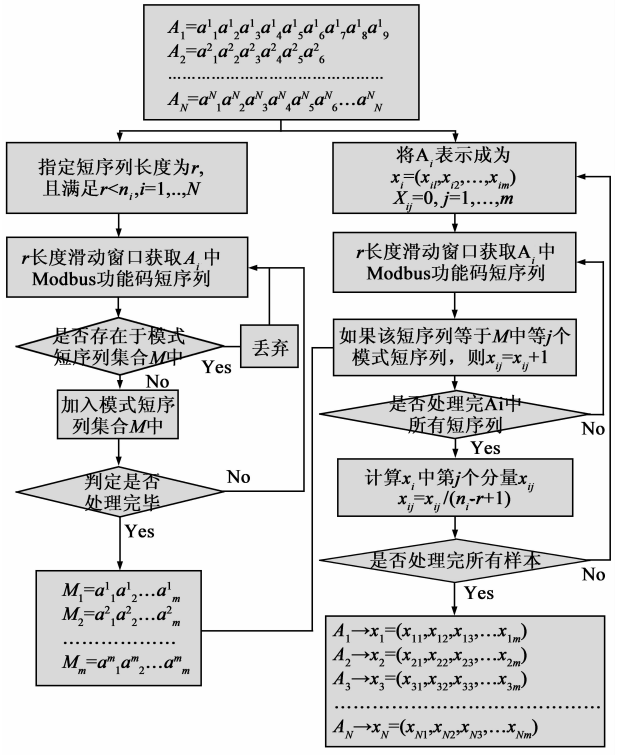

图2 数据预处理流程

Step1 指定短序列长度 r.对于 Modbus 功能码序列样本集合$\{A_n\}$,$n=1,\cdots,N$,其中 N 为样本数目,$A_i=a^i_1a^i_2\cdots a^i_{n_i}$,$a^i_j\in\{$Modbus 功能码$\}$,$j=1,\cdots,n_i$,$n_i$ 表示第 i 个样本中含有的 Modbus 功能码数量,需要满足 $j=1,\cdots,n_i$,n_i 表示第 i 个样本中含有的 Modbus 功能码数量,需要满足 $r\leqslant n_i,\forall\ i=1,\cdots,N$.

Step2 获取模式短序列集合 M.按照长度为 r 的滑动窗口循环处理 Modbus 功能码样本集合中的每一个元素 A_i,至多可以得到 $\sum_{i=1}^{N}n_i-rN+N$ 个包含 r 个 Modbus 功能码的序列,去除其中重复的成分并且仅保留一份可以得到模式短序列集合 $M=\{M_1,M_2,\cdots,M_m\}$,$m\leqslant\sum_{i=1}^{N}n_i-rN+N$.

Step3 Modbus 功能码序列的向量表达.将任意的 Modbus 功能码序列 $A_i=a^i_1a^i_2\cdots a^i_{n_i}$,按照每一个模式短序列出现的频率,构造成 SVM 特征向量 $\boldsymbol{x}_i=(x_{i1},x_{i2},\cdots,x_{im})$,其中 $\boldsymbol{x}_i$ 的第 j 个分量 x_{ij} 表示短序列模式集合 M 中的第 j 个分量 m_j 出现的频率,计算公式为 $x_{ij}=g(m_j)/(n_i-r+1)$,式中 $g(m_j)$ 表示在 A_i 中模式短序列 m_j 出现的次数.

显然,使用该方法对 Modbus TCP 流量进行预处理有以下优点:首先,可以将含有不同数目的数据包流量映射为相同长度的向量,使得样本更加灵活;其次,用该方法得到向量的每一维的数值表示了对应的模式短序列出现的频率特性,使得该向量对序列的描述更加合理.

4 PSO-SVM 异常检测模型

异常检测的方法主要有数学统计、特征选择、神经网络、机器学习、数据挖掘等,但是这些方法几乎都需要大量的训练样本,而在样本较少的情况下会产生较大的“误报率”和“漏报率”.

与上述方法相比,SVM 算法具有如下优势:首先,SVM 适合在有限样本情况下获得最优解,本文在所搭建的实验环境下捕获的数据恰好是小容量样本的情况;其次,SVM 算法最终演化为凸二次规划问题,可避免神经网络中的局部极值问题;再次,可将非线性可分问题从低维空间变换到高维空间,从而实现线性可分,与此同时引入核函数,避免了维数灾难的问题,使得算法复杂度与样本的空间维数无关[15,16].

在实际应用中,SVM 的性能很大程度上要受到惩罚因子 C 和核函数类型及参数选择等因素制约,设计者更多是根据个人经验进行选择和设置.粒子群算法 PSO 是一种群体智能算法,广泛应用于参数优化,能够有效提高搜索效率[17].

本文设计了一种基于 PSO-SVM 对 Modbus 功能码序列进行异常检测的模型,算法流程如图 3 所示.该算法利用 PSO 算法对 SVM 模型进行参数寻优与结构优化,从而提升了模型分类识别精度.

PSO 算法首先在可行解的空间初始一组粒子,每个粒子代表问题的一个潜在最优解,用位置、速度、适应度值表示粒子特征,特征值需要根据实际问题设置最优值计算函数,该值可以表示粒子的优劣.粒子在解空间中运动,每一次进化需要更新个体极值、群体极值、位置、速度等特征.其中,个体极值代表个体所经历的

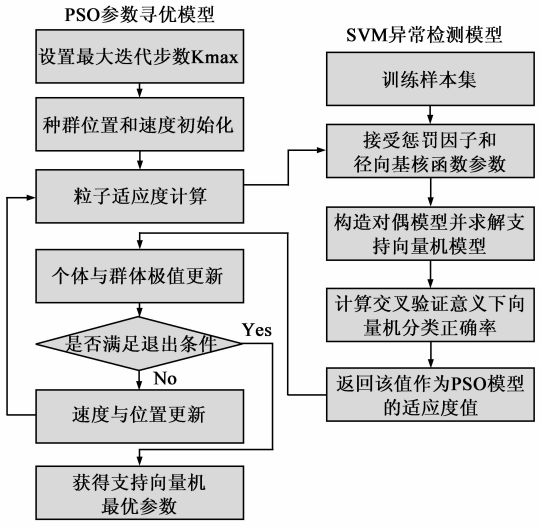

图3　PSO-SVM异常检测模型

位置中适应度值最优的位置,群体极值是指所有粒子在经历的位置中适应度值最优的位置.粒子群算法 PSO 进行速度与位置的公式如下[18,19]:

$$\boldsymbol{V}^{k+1} = \omega \boldsymbol{V}^{k} + c_1 r_1 (\boldsymbol{P}^{k} - \boldsymbol{X}^{k}) + c_2 r_2 (\boldsymbol{G}^{k} - \boldsymbol{X}^{k}) \quad (1)$$

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^{k} + \boldsymbol{V}^{k+1} \quad (2)$$

上式中角标 k 与 $k+1$ 分别表示上一轮迭代和本轮的属性,$\boldsymbol{V}$ 表示速度,$\boldsymbol{P}$ 表示个体极值,$\boldsymbol{G}$ 表示群体极值,$\boldsymbol{X}$ 表示位置.惯性因子 c_1 和 c_2 为非负常数,加速度因子 r_1 和 r_2 为 0 到 1 之间的随机数,基于 PSO 基本思想对 SVM 参数惩罚因子 C 以及径向基核函数 σ 寻优的算法流程如下:

Step0　设置最大迭代步数 $k_{\max}$.

Step1　随机生成种群以及相关参数初始化.随机生成种群位置 $X = (\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N)$,$N$ 为粒子数目,其中 $\boldsymbol{X}_i = (x_{ic}, x_{i\sigma})$表示第 i 个粒子由两个分量构成,分别代表 SVM 惩罚因子 C 和径向基核函数 σ 的位置,随机生成对应每一个位置均有速度 $\boldsymbol{V}_i = (V_{ic}, V_{i\sigma})$.设置位置两个分量的限定范围是$[X_{c\min}, X_{c\max}]$和$[X_{\sigma\min}, X_{\sigma\max}]$.

Step2　进行粒子适应度 $F(\boldsymbol{X}_i)$计算.粒子适应度值 $F(\boldsymbol{X}_i)$选取以分量 x_{ic} 和 $x_{i\sigma}$ 为参数的基于 SVM 的 Modbus 功能码序列检测的交叉验证意义下的分类正确率.

Step3　根据适应度值更新个体极值及群体极值.如果适应度值 $F(\boldsymbol{X}_i^k) > F(\boldsymbol{X}_i^{k-1})$,$\boldsymbol{P}^k = X^k$,否则 $\boldsymbol{P}^k = \boldsymbol{X}^{k-1}$.如果存在 j 使得 $F(\boldsymbol{X}_j^k) > F(\boldsymbol{X}_i^k)$均成立,且 $F(\boldsymbol{X}_j^k) > F(\boldsymbol{G}^{k-1})$,则另 $\boldsymbol{G}^k = \boldsymbol{X}_j^{k-1}$,否则 $\boldsymbol{G}^k = \boldsymbol{G}^{k-1}$.

Step4　判断是否满足退出迭代条件.如果超过迭代次数极值或连续 50 次适应度值的变化没有超过 0.01%,则退出迭代过程,并且此时的群体极值 $\boldsymbol{G}$ 即为所要求的最佳参数.

Step5　按照粒子速度与位置更新公式进行更新.每一轮更新结束后需要判定位置各维是否限定在规定范围$[X_{c\min}, X_{c\max}]$和$[X_{\sigma\min}, X_{\sigma\max}]$内,对于超过范围的分量需要限定在该范围之内,例如:如果 $x_{ic} < x_{c\min}$,则设置 $x_{ic} = x_{c\min}$;如果 $x_{ic} > x_{c\max}$,则 $x_{ic} = x_{c\max}$.

在 Step2 中的粒子适应度计算函数 $F(\boldsymbol{X})$选择 SVM 的交叉验证分类正确率.基于 SVM 的 Modbus 功能码异常检测建立模型的步骤如下:

Step1　接受 PSO 参数优化流程传递的惩罚因子 C 和径向基核函数参数 σ.

Step2　赋予所有样本类别标签.正常功能码序列样本标签设置为 1,异常功能码序列样本标签设置为 -1.

Step3　构造对偶求解支持向量机模型[20].

$$\min_{\alpha} Q = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{n}\alpha_i \quad (3)$$

$$\text{s.t.} \quad \sum_{i=1}^{n}\alpha_i y_i = 0, 0 \leqslant \alpha_i \leqslant C, i = 1, \cdots, n \quad (4)$$

得解 $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*)$.

Step4　构造决策函数[20].

$$b^* = y_j - \sum_{i=1}^{n} y_i \alpha_i^* K(\boldsymbol{x}_i, \boldsymbol{x}_j), j \in \{j \mid 0 < \alpha_j^* < C\} \quad (5)$$

$$f(x) = \operatorname{sgn}\Big(\sum_{i=1}^{n}\alpha_i^* y_i^* K(\boldsymbol{x}_i, \boldsymbol{x}) + b^*\Big) \quad (6)$$

Step5　根据判别函数计算 SVM 的分类准确率,且将该值返回 PSO 参数优化流程,作为粒子适应度计算函数 $F(X)$的取值.

5　仿真实验结果对比与分析

5.1　实验数据获取

目前,国内外尚无公开的工业控制系统 Modbus TCP 通讯的数据集可供测试评估.因此为了进行 Modbus 功能码序列异常检测的可行性分析和各种算法的验证,本文设计并搭建了仿真实验环境,拓扑结构如图 4 所示.

该仿真环境控制单元层选择施耐德 M340PLC,CPU 型号为 2020,与上位机之间的通讯采取 Modbus TCP 协议.数据采集层选取 KingSCADA 软件开发监控画面,工程师站选取 UnityPro 对 PLC 进行组态编程,攻击模拟源模拟遭受病毒感染的站点对重要控制器 PLC 发送恶意流量.该环境模拟了液体容器中液位的控制,相关阀门的开闭,液位传感器的数值均在 PLC 中进行逻辑上的模拟.

系统运行时,抓取网络中的 Modbus TCP 通讯流量,剔除 TCP 机制中的三次握手、确认重传等不包含有

Modbus 功能码数据包,则得到 Modbus TCP 客户端和 Modbus TCP 服务器端的通讯流量,因为协议规定响应帧和请求帧中 Modbus 功能码相同,因此可以进一步剔除所有从服务器端回应客户端的通讯流量,而剩余的客户端到服务器端的通讯流量恰好是保障工业控制系统中重要控制器的关键.

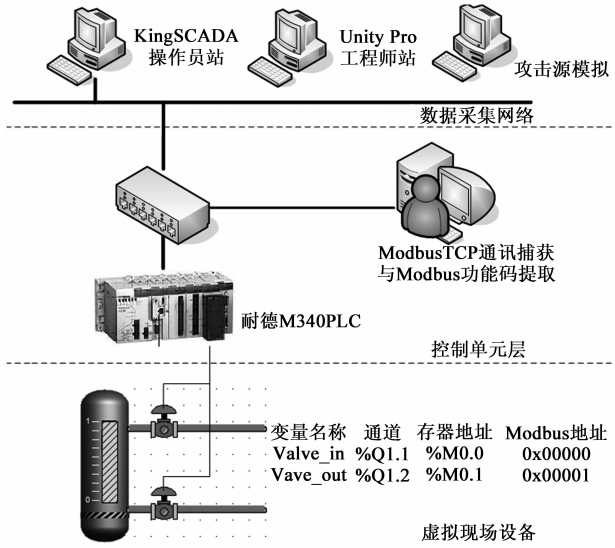

图4 工业控制系统仿真实验环境

5.2 数据分析

在上文中搭建的工业控制系统仿真实验环境中获取了 114 个 Modbus 功能码序列,选择短序列长度 $r=3$,经过预处理后得到人工智能方法可处理的数据,其中含有 64 个 +1 类样本和 50 个 −1 类样本.

PSO 算法的参数 c_1 和 c_2 的参数选择方法已经较为成熟,其和一般不超过 4,其选择体现了设计者对于粒子群中个体知识和群体知识的权衡,这里根据经验设置 PSO-SVM 模型中 $c_1=1.5$, $c_2=1.7$,种群大小设置为 20,进化代数设置为 100,采用 5 折交叉校验方式求取 SVM 分类准确率.同时,也分别设计了基于网格化参数寻优[21]的 SVM、标准 RBF 神经网络与 BP 神经网络的异常检测模型,进行验证对比.

PSO-SVM 的异常检测效果如图 5、6、7 所示,显然在粒子群迭代寻优过程中,适应度值可以较快的收敛,说明 PSO 寻优过程效率较高.该模型的测试集分类准确率为 95.83%,效果良好.训练集分类准确率为 100%,验证了设计的支持向量机模型具有较强的学习能力.总之,PSO-SVM 模型兼具优化过程速度快和 SVM 泛化能力强的特点.

为使实验数据能够更加客观真实反映模型性能,每种模型做了 30 次的分类实验,得到关键指标的统计数据如表 1 所示.

对比表 1 可知,由于样本数目较少,各种模型的训练集分类准确率均能达到 100%,然而两种 SVM 方法在测试集合上的分类效果要显著优于 RBF 神经网络与 BP 神经网络,这也说明了 SVM 具有较强的泛化能力.

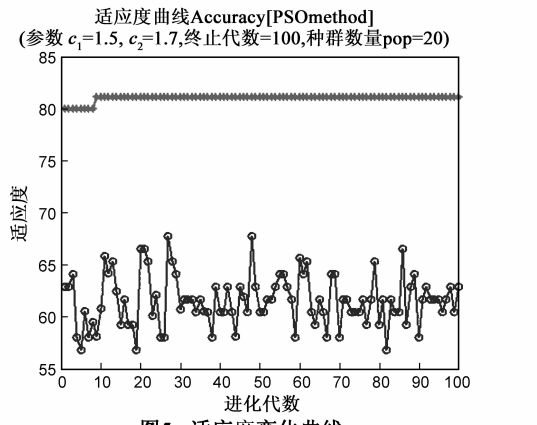

图5 适应度变化曲线

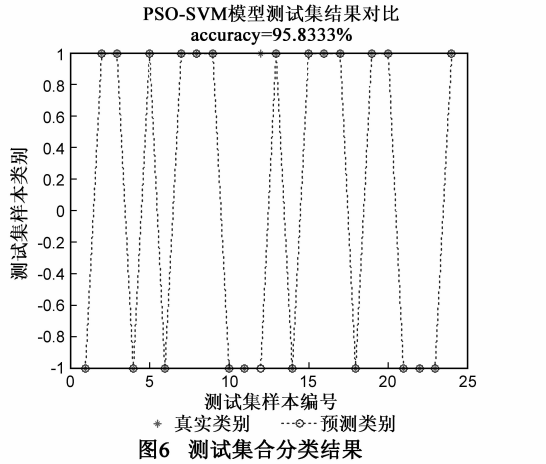

图6 测试集合分类结果

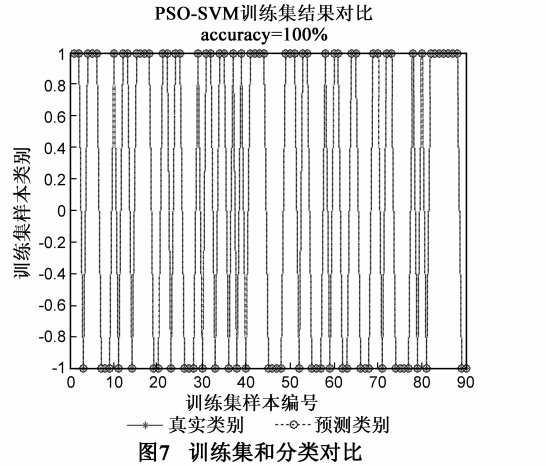

图7 训练集和分类对比

对于两种支持向量机而言,PSO-SVM 模型的测试集分类能力要略优于网格化寻优的 C-SVM 支持向量机模型,搜索到的惩罚因子也远小于后者,这意味着该模型具有更少的支持向量,耗费更少的存储空间来描述分类模型,形式上更加简洁.由于网格化寻优过程实际

上是在 C 与 σ 形成的二维空间上进行最佳参数的搜索,实验过程中每个维度上均等地分为 100 点,因此该算法事实上调用了 100 * 100 = 10000 次 C-SVM 过程,而 PSO 搜索过程设置了迭代步数上限为 100,粒子群数目为 20,因此 PSO-SVM 模型共调用 100 * 20 = 2000 次 C-SVM 过程,因此所耗费的时间为网格化参数寻优过程的 1/5.表 1 中时间消耗之比约为 0.17,说明效率至少提高了 4 倍.事实上多次实验结果均类似于图 5,粒子群的最佳适应度曲线在迭代步数在 10 左右即达到收敛状态而不再变化,因此在最佳状态下,与网格化的 C-SVM 的搜索效率之比,可以达到(10 * 20)/(100 * 100) = 1/40,C-SVM 的时间复杂度大概为 $O(N^2)$,因此本文设计的 PSO-SVM 模型的时间复杂度应为 200 * $O(N^2)$,远低于网格化 C-SVM 参数寻优方法的时间代价.

表 1　四种模型结果对比

类别	PSO-SVM	C-SVM	RBF	BP
C_{best}	1.22627	2.75197	/	/
σ_{best}	0.1	0.04	/	/
ACC	83.54%	91.85%	/	/
$ACC_{训练集}$	100%	100%	100%	100%
$ACC_{测试集}$	84.17%	79.58%	50%	54%
Time	23.68s	139.83s	1.8s	4s

此外,为了进一步说明参数选择对于结果的影响,又做了一组验证性试验.该实验将表 1 中的 PSO-SVM 模型的最优参数 C_{best}和 σ_{best}指定为 C-SVM 的相关参数,并且随机选择几组不同的 C 和 σ 进行试验对比,每组参数进行 10 次试验得到关键指标的平均数值如表 2 所示.

表 2　参数选择对 SVM 分类效果影响展示

类别	第一组	第二组	第三组	第四组	第五组	第六组
C	1.22627	2.75197	5	10	0.1	0.01
σ	0.1	0.04	1	2	0.1	0.1
ACC_{cv}	86.24%	89.34	58.17%	55.77%	55.57%	55.99%
$ACC_{训练集}$	100%	100%	100%	100%	100%	100%
$ACC_{测试集}$	88.32%	83.03%	64.34%	61.33%	58.22%	56.66%

显然,参数优化的过程对于支持向量机的性能具有很大的影响,这也充分说明了 PSO 算法寻找 SVM 参数的合理性和有效性.由于 PSO-SVM 模型具有更小的惩罚因子 C,因而理论上获得较少的支持向量.根据式(6)可知,支持向量机模型进行判别所需要的时间和支持向量的数量成正比,所以在实时检测方面的性能要优于网格化寻优的 SVM 方法.但是由于本文的样本数量较少,而支持向量的多少与样本数目相关,因而 PSO-SVM 与网格化 SVM 方法的实时性能对比的结果不会太明显,在 Matlab 中对上述所有样本判别 1000 次的总时间为 4.2s,即每个样本大概需要花费 3.5×10^{-5}S,说明算法运行时间能够满足工业控制系统的实时性要求.

6　结论

针对攻击者利用 Modbus TCP 协议应用层设计缺陷和防火墙与入侵检测系统安全规则与"白名单"策略发动的攻击,无法被识别或者拦截的问题,本文提出了重点分析 Modbus 功能码字段,将 Modbus TCP 通讯流量转换为异常检测模型所需数据形式的预处理方法,设计了一种利用粒子群算法 PSO 进行参数寻优的 PSO-SVM 模型.该模型根据 Modbus 功能码序列中的模式短序列出现的频率,进行 Modbus TCP 通讯流量的异常检测,取得了良好的效果.

本文提出模型与算法在实时性方面还相对薄弱,未来的研究工作可以将工业通讯协议中几个不同字段作为 Modbus 通讯流量特征,并且在数据降维处理方面展开数据预处理的工作,以得到既能有效表示通讯行为特征又能保持数据的精简.

参考文献

[1] 魏钦志.工业网络控制系统的安全与管理[J].测控技术,2013,32(2):87 - 92.
Wei Qin-zhi. Industrial network control system security and management[J]. Measurement & Control Technology, 2013, 32(2):87 - 92.(in Chinese)

[2] 彭勇,江长青,谢丰,等.工业控制系统信息安全研究进展[J].清华大学学报(自然科学版),2012,52(10):1396 - 1405.
Peng Yong, Jiang Changqing, Xie Feng, et al. Industrial control system cyber security research[J]. Journal of Tsinghua University(Science and Technology), 2012, 52(10):1396 - 1405.(in Chinese)

[3] 熊琦,竞小伟,詹峰.美国石油天然气行业 ICS 系统信息安全工作综述及对我国的启示[J].中国信息安全,2012,27(03):80 - 83.
Xiong Qi, Jing Xiaowei, Zhan Feng. Summary and implications for China of the information security work of the ICS system in the oil and gas industry in America[J]. China Information Security, 2012, 27(3):80 - 83.(in Chinese)

[4] 2013 工业控制系统及其安全性研究报告[EB/OL]. http://www.nsfocus.com/4_research/4_6.html, 2013-06-24.

[5] 孙大林,蒋大明.Modbus/TCP 的安全性及其在工业监控系统中的应用[J].中国安全生产科学技术,2006,2(2):92 - 95.
Sun Da-lin, Jiang Da-ming. Modbus/TCP protocol safety and its application in industrial monitoring and control system[J]. Journal of Safety Science and Technology, 2006, 2(2):92 - 95.(in Chinese)

[6] 王婷婷.SCADA 系统中数据传输安全性研究[D].上海:华东理工大学,2012.
Wang Tingting. Security research on SCADA system data transmission[D]. Shanghai: East China University of Science and Technology,2012.(in Chinese)

[7] 张云贵,赵华,王丽娜.基于工业控制模型的非参数 CUSUM 入侵检测方法[J].东南大学学报(自然科学版),2012,42(S1):55 – 59.
Zhang Yungui, Zhao Hua, Wang Lina. A non-parametric CUSUM intrusion detection method based on industrial control model[J]. Journal of Southeast University (Natural Science Edition),2012,42(S1):55 – 59.(in Chinese)

[8] Javier J. Using SNORT for intrusion detection in MODBUSTCP/IP communications[EB/OL]. http://www.giac.org/paper/gcia/7218/snort-intrusion-detection-modbus-tcp-ip-communications/124438,2013-08-30.

[9] Venkat P, Matthew F. Transparent Modbus TCP Filtering with Linux[EB/OL]. http://modbusfw.Sourceforge.net,2013 – 08 – 30..

[10] 夏春明,刘涛,王华忠,等.工业控制系统信息安全现状及发展趋势[J].信息安全与技术,2013,2:13 – 17.
Xia Chun-ming, Liu Tao, Wang Hua-zhong, et al. Industrial control system security analysis[J]. Information Security and Technology,2013,2:13 – 17.(in Chinese)

[11] GB/T 19582.1-2008.基于 Modbus 协议的工业自动化网络规范第 1 部分:Modbus 应用协议[S].

[12] GB/T 19582.3-2008.基于 Modbus 协议的工业自动化网络规范第 3 部分:Modbus 协议在 TCP/IP 上的实现指南[S].

[13] Eric K. Industrial Network Security securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control System[M]. Syngress,2011.

[14] Peter H Rodrigo C, Mauricio P, et al. Attack taxonomies for the Modbus protocols[J]. Critical Infrastructure protection, 2008,(1):37 – 44.

[15] 李昆仑,赵俊忠,黄厚宽,田盛丰.基于 SVM 技术的入侵检测[J].信息与控制,2003,32(6):495 – 499.
Li Kun-lun, Zhao Jun-zhong, Huang Hou-kuan, et al. An intrusion detection method based on SVM[J]. Information and Control,2003,32(6):495 – 499.(in Chinese)

[16] 邬俊,鲁明羽,刘闯.基于混合学习框架的 SVM 反馈算法研究[J].电子学报,2010,38(9):2101 – 2106.
Wu Jun, Lu Ming-yu, Liu Chuang. SVM-based scheme within hybrid learning framework for image retrieval[J]. Acta Electronica Sinica,2010,38(9):2101 – 2106.(in Chinese)

[17] 陈国初,俞金寿.微粒群优化算法[J].信息与控制,2005,34(3):318 – 323.
Chen Guo-chu, Yu Jin-shou. Particle swarm optimization algorithm[J]. Information and Control, 2005, 34(3): 318 – 323.(in Chinese)

[18] Jiang B, Wang N, Wang LP. Particle swarm optimization with age-group topology for multimodal functions and data clustering[J]. Communications in Nonlinear Science and Numerical Simulation,2013,18(11):3134 – 3145.

[19] Cabrerizo FJ, Herrera-Viedma E, Pedrycz W. A method based on PSO and granular computing of linguistic information to solve group decision making problems defined in heterogeneous contexts[J]. European Journal of Operational Research, 2013,230(3):624 – 633.

[20] 张学工.关于统计学习理论与支持向量机[J].自动化学报,2000,26(1):32 – 40.
Zhang Xuegong. Introduction to statistical learning theory and support vector machine[J]. Acta Automation,2000,26(1):32 – 40.(in Chinese)

[21] 李林,张晓龙.基于 RBF 核的 SVM 学习算法的优化计算[J].计算机工程与应用,2006(29):190 – 204.
Li Lin, Zhang Xiao-long. Optimization of SVM with RBF kernel[J]. Computer Engineering and Applications, 2006(29): 190 – 204.(in Chinese)

作者简介

尚文利 男,1974 年生于黑龙江省北安市.博士,副研究员,硕士生导师.主要研究方向为工业控制系统信息安全、嵌入式系统、机器学习等.现为中国科学院沈阳自动化研究所工业控制系统信息安全方向学术带头人.
E-mail:shangwl@sia.cn

张盛山 男,1988 年生于辽宁省瓦房店市.中国科学院大学硕士研究生.研究方向为嵌入式系统、工业控制系统信息安全.
E-mail:zhangshengshan@sia.cn