文章编号:1671-4598(2024)01-0008-08 DOI:10.16526/j.cnki.11-4762/tp.2024.01.002 中图分类号:TP399 文献标识码:A

基于 PSO-KM 聚类分析的通信网络恶意攻击代码检测方法

李 梅,朱明宇
(苏州高博软件技术职业学院 信息与软件学院,江苏 苏州 215163)

摘要:恶意代码的快速发展严重影响到网络信息安全,传统恶意代码检测方法对网络行为特征划分不明确,导致恶意攻击代码的识别率低、误报率高,研究基于 PSO-KM 聚类分析的通信网络恶意攻击代码检测方法;分析通信网络中恶意攻击代码的具体内容,从网络层流动轨迹入手提取网络行为,在 MFAB-NB 框架内确定行为特征;通过归一化算法选择初始处理中心,将分类的通信网络行为特征进行归一化处理,判断攻击速度和位置;实时跟进通信网络数据传输全过程,应用适应度函数寻求恶意代码更新最优解;基于 PSO-KM 聚类分析技术构建恶意代码数据特征集合,利用小批量计算方式分配特征聚类权重,以加权平均值作为分配依据检测恶意攻击代码,实现检测方法设计;实验结果表明:在文章方法应用下对恶意攻击代码检测的识别率达到 95.0%以上,最高值接近 99.7%,误报率可以控制在 0.4%之内,具有应用价值。

关键词:恶意攻击代码;通信网络;PSO-KM 聚类分析;聚类权重;网络行为特征;行为优劣程度

Detection Method of Malicious Attack Codes in Communication Network Based on PSO-KM Cluster Analysis

LI Mei, ZHU Mingyu
(School of Information and Software, Suzhou Global Institute of Software Technology, Suzhou 215163, China)

Abstract: The rapid development of malicious code has seriously affected network information security. Traditional malicious code detection methods do not clearly divide network behavior characteristics, causing the low recognition rate and high false positive rate of malicious attack code. Therefore, a malicious attack code detection method for communication network based on PSO-KM clustering analysis is researched. The specific content of malicious attack code in communication network is analyzed, and the network behavior is extracted from the flow trajectory of network layer, and the behavior characteristics are determined in the MFAB-NB framework. The initial processing center is selected by the normalization algorithm, and the behavior characteristics of the classified communication network are normalized to judge the attack speed and location. The whole process of communication network data transmission is followed up in real time, and the fitness function is applied to seek the updating optimal solution of malicious code. The feature set of malicious code data is constructed based on the PSO-KM clustering analysis technology, and the small batch calculation method is used to allocate the weight of the feature cluster. The weighted average value is used as the distribution basis to detect the malicious attack code, realize the design of detection method. The experimental results show that under the application of this method, the correct recognition rate of malicious attack code detection can reach more than 95.0%, the highest value is close to 99.7%, and the false positive rate can be controlled within 0.4%, and it has better application value.

Keywords: malicious attack code; communication network; PSO-KM clustering analysis; cluster weight; network behavior characteristics; degree of good or bad behavior

0 引言

在信息技术飞速发展的势态下,以互联网为代表的信息网络发生了巨大变化,直接影响人们的生产和生活习惯。与此同时微电子和传感器以及无线通信技术的快速发展,也不断地推动互联网的迅速发展,从而产生了严重的网络安全问题。由于通信网络中包含有大量的数据节点,在数据存储和数据处理中容易受到恶意节点的攻击,当通信网络的性能减低时会引发网络崩溃[1],造成通信网络的大面积瘫痪,因此通信网络的运行需要在极大地保障下完成工作,通信网络恶意攻击代码检测方法对恶意攻击代码的检测和防御研究具有重要意义。国内外众多相关领域的专家非常注重恶意代码检测的研究,现阶段恶意攻击代码的分析和检测已经成为被广泛讨论的话题,并从不同的角度给出了多种应对策略,为实现通信网络的安全运行提供了多种技术支持。传统通信网络安全问题研究中通常使用

收稿日期:2023-02-07; **修回日期**:2023-04-17。
基金项目:江苏省高等职业教育高水平专业群(苏教职函[2021]1号);江苏省高等职业教育高水平骨干专业建设项目(苏教高[2017]17号)。
作者简介:李 梅(1981-),女,硕士,讲师。
引用格式:李 梅,朱明宇.基于 PSO-KM 聚类分析的通信网络恶意攻击代码检测方法[J].计算机测量与控制,2024,32(1):8-15.

硬件进行入侵防御，文献［2］研究了基于被动分簇算法的即时通信网络协议漏洞检测，其引入了被动分簇算法，采用该算法中先声明者优先机制挑选簇首，结合均衡原则明确网关节点，并且结合前向反馈网络和支持向量机，构建通信网络协议漏洞检测方法，实现检测。但对于通信网络的内部攻击没有任何防御效果，且不能适用于通信网络的动态变化，其正确识别率较低，无法保障通信网络安全。文献［3］研究了一种基于多特征集成学习的恶意代码静态检测框架，该方法通过提取恶意软件的非 PE 结构等特征，构建特征相匹配模型，通过集成算法提升模型稳定性，实现恶意代码检测。但也只能对一部分数据进行保护，无法完全辨别不同类型的恶意代码，恶意代码正确识别率低。

针对上述方法存在的问题，本文选择 PSO-KM 聚类分析技术作为支撑，借助该技术的动态特征提取优势确定网络中是否存在恶意入侵，设计通信网络的恶意攻击代码检测方法，以期通过该检测方法提高通信网络的安全性。

1　确定通信网络流动轨迹中恶意攻击行为特征

随着通信能力的逐渐增强网络中各个节点的功能发生巨大转变，通信节点从单纯的信息采集扩展到网络系统的数据处理以及存储等更多任务，因此用于通信网络系统数据处理的节点经常受到恶意攻击引起网络故障。为保证数据不被窃取和篡改以及丢失等要求，需要对通信网络数据的流动轨迹的行为特征进行分析，以无线传感器节点为通信网络的数据处理主模块，主要监测通信网络数据流转区域内的数据采集和转换，对其基本体系结构进行分析，如图 1 所示。

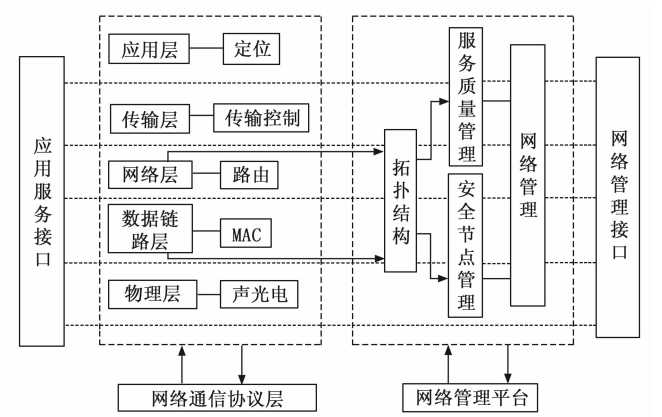

图 1　通信网络基本体系结构

根据图 1 内容所示，通信网络中的无线传感器体系结构可以分为网络通信协议、网络数据管理和应用支撑 3 个部分，其中最主要的网络通信协议层级，该层包含数据的链路层和传输层，由于数据链路和传输在该层，则此处是恶意攻击重点关注位置，因此容易产生恶意攻击行为[4]。对其进行具体功能划分：网络通信协议中物理层，其主要借助无线电和红外线等传输媒介，完成节点信号的调制和数据的收发。数据链路层实现数据信息的成帧检测以及错误控制，网络层是对路由的控制和维护，传输层用来控制通信数据信息流的传输，应用层则是连接网络平台中的对应软件。

针对通信网络的基本体系结构，将其放置于在 MFAB-NB 框架内具体分析恶意攻击的具体类型，可以大体上划分为：对网络协议层的攻击、对机密性和认证性数据的攻击、以及对数据服务完整性的攻击，由于网络协议层为主要的数据流动轨迹，直接对网络协议层进行分析，将其受到的恶意攻击行为特征进行类型划分，按照物理层恶意攻击、链路层恶意攻击、网络层恶意攻击以及传输层恶意攻击进行展示，具体如表 1 所示[5]。

表 1　网络协议层中流动轨迹恶意攻击行为特征分类

网络层级	攻击手段	具体类型	后果
物理层	拥塞攻击	1. 持续型 2. 欺骗型 3. 随机型 4. 交互型	1. 占用信道 2. 数据错误 3. 破坏数据 4. 数据碰撞
	物理破坏	1. 改动位置 2. 内部攻击	1. 占用节点 2. 窃取数据
链路层	碰撞攻击	1. 引发冲突 2. 不间断攻击	1. 丢失节点 2. 消耗资源
	耗尽攻击	1. 漏洞攻击 2. 虚假协议	1. 消耗资源 2. 更改协议
	不公平竞争	1. 优先占领 2. 融合节点	1. 占据信道 2. 节点死亡
网络层	路由欺骗	1. 篡改路由 2. 欺诈路由 3. 重复路由	1. 虚假链路 2. 缩减寿命 3. 浪费资源
	转发攻击	1. 随机转发 2. 合并数据包	1. 丢弃数据 2. 终端消失
	黑洞攻击	1. 散播节点 2. 欺骗节点 3. 配合攻击	1. 链路造假 2. 破坏传输 3. 真假融合
	女巫攻击	1. 伪造 ID 2. 破坏机制	1. 盗用节点 2. 误导节点
	虫洞攻击	1. 配合攻击 2. 延时链路	1. 吸引节点 2. 消耗资源
传输层	洪泛攻击	1. 包围节点 2. 扩大距离	1. 节点误判 2. 消耗资源
	失步攻击	1. 破坏链路 2. 欺骗节点	1. 消耗资源 2. 节点死亡

根据表中内容所示对不同的层级攻击行为进行分析，如以物理层来讲拥塞攻击主要是在通信网络链路中不间断的发送干扰信号，造成通信节点不能正常进行数据传输从而破坏网络，物理攻击则是将正常节点进行伪装，从正常节点的位置中发起攻击对网络安全造成威胁[6]。其他协议层中的攻击手段也不尽相同，基本上都是在链路中直接对节点产生攻击，消耗网络通信量的基础上造成正常节点的

死亡，以此攻击方能够在数据融合过程中导入虚假数据。在此基础上选择归一化算法对特征归类，计算不同恶意攻击行为的特征值，判断链路中出现攻击行为的可能性。

2 归一化处理通信网络恶意攻击行为特征

无论哪一种通信网络攻击均含有多重身份 ID，其 ID 数据可以是伪造的，也可以直接盗用正常节点，一旦恶意攻击代码形成，则通信网络中的数据运行机制会遭受严重破坏，对数据的存储和处理以及分配均会产生影响[7]。在有限条件下对恶意攻击代码进行特征计算，将恶意攻击行为进行归一化处理，通过恶意攻击代码的接受信号强度指标，对不同的攻击行为进行信道监测[8]。信号强度指标能够比较节点接收数据的功率大小，通过功率比值对信号源节点进行监测，计算节点中是否存在大于其自身的能力优势。利用普通节点 m 和 n 进行比较，设定被恶意攻击代码选择的节点为簇头节点 b，则：

$$\frac{RSSI_m}{RSSI_n}=\frac{RSSI_b\times\frac{\upsilon}{c_m^{\chi}}}{RSSI_b\times\frac{\upsilon}{c_n^{\chi}}}=\left(\frac{c_m}{c_n}\right)^{\chi} \tag{1}$$

公式中：节点 m 的接收信号强度为 $RSSI_m$。节点 n 的接收信号强度为 $RSSI_n$。簇头节点 b 的接收信号强度为 $RSSI_b$[9]。阈值为 υ。强度比例为 χ。簇头距离为 c。以公式可知恶意攻击代码的节点选择原则，只与其到簇头的距离相关。假定恶意节点中存在良好总身份 ID，分别为 x_1 和 x_2，与其相关的存在有 4 组普通节点为 z_1、z_2、z_3、z_4，则形成判断公式如下：

$$\begin{cases}\dfrac{RSSI_{z_1}^{x_1}}{RSSI_{z_2}^{x_1}}=\dfrac{RSSI_{z_1}^{x_2}}{RSSI_{z_2}^{x_2}}\\ \dfrac{RSSI_{z_1}^{x_1}}{RSSI_{z_3}^{x_1}}=\dfrac{RSSI_{z_1}^{x_2}}{RSSI_{z_3}^{x_2}}\\ \dfrac{RSSI_{z_1}^{x_1}}{RSSI_{z_4}^{x_1}}=\dfrac{RSSI_{z_1}^{x_2}}{RSSI_{z_4}^{x_2}}\end{cases} \tag{2}$$

若公式中 x_1 和 x_2 到 z_1、z_2、z_3、z_4 4 组节点的距离相等，则可以得出 x_1 和 x_2 同时处于同一物理位置，即 x_1 和 x_2 为同一个节点，但由于身份 ID 不同，则说明在该节点处存在有恶意攻击代码[10]，并且不同的信号强度下消耗节点资源也不同，一般情况下，恶意攻击的主要目的是消耗服务器的资源。在完成节点位置初步确定后，为进一步提高位置确定的准确性，以归一化处理方式假定攻击原理，设定不同的攻击行为特征进行初始位置，$k_l=\{k_{l1},k_{l2},\ldots,k_{lj}\}$ 表示为代码初始位置，l 表示初始位置，$h_l=\{h_{l1},h_{l2},\ldots,h_{lj}\}$ 表示初始攻击速度，代码的个体极值可表示为 $g_f=\{g_{f1},g_{f2},\ldots,g_{fj}\}$，能够被找到的全局极值表示为 $g_d=\{g_{d1},g_{d2},\ldots,g_{dj}\}$[11]。结合 PSO 算法，确定代码恶意攻击行为特征的攻击速度以及位置，计算公式如下：

$$h_{ls}(a)=\delta h_{ls}(a-1)\beta_1 rand(a)(g_f(a)-k_{ls}(a))+\beta_2 rand(a)(g_f(a)-k_{ls}(a)) \tag{3}$$

$$k_{ls}(a)=k_{ls}(a-1)+h_{ls}(a) \tag{4}$$

上式种（3）为恶意攻击代码的速度更新方式，公式（4）表示恶意攻击代码位置的更新方式[12]。其中 $s=1,2,\ldots,j$、$l=1,2,\ldots,p$，p 为攻击代码数量。δ 表示攻击代码飞行过程中的惯性系数，当 δ 越大时表示攻击代码的运行速度就越大，产生的破坏能力就越强，反之当 δ 值越小说明其搜索能力越弱，并且 δ 可以为定值也可以为变值。系数 β_1 和 β_2 表示攻击代码的学习因子，学习因子描述代码对整个攻击过程的认知程度，在 β_1 取值较小时，即 $\beta_1\leqslant 0.1$，代码会偏移向攻击区域，β_2 较大时，即 $\beta_2\geqslant 1.0$，代码可以迅速到达指定目标区域。$rand(a)$ 表示随机函数，a 表示 (0,1) 之间的数值[13]。$k_{ls}(a)$ 表示攻击代码目前的位置。$k_{ls}(a-1)$ 表示历史攻击中寻找到的最优位置。$h_{ls}(a)$ 表示恶意攻击代码目前的攻击速度。

至此完成归一化处理通信网络恶意攻击行为特征，确定了通信网络恶意攻击行为的攻击速度和位置。但是在多个变量的影响下攻击代码会逐渐靠近正常节点，此时，需要采用适用度函数计算其最优解，预先判断其最佳攻击位置。

3 粒子群优化算法寻找恶意攻击代码更新最优解

在第 2 章节通过归一化处理通信网络恶攻击意行为特征，获取高质量特征数据后，引入粒子群优化算法（PSO），寻找恶意代码更新最优解。在获取攻击节点与正常节点距离关系的基础上，分类不同攻击行为，由攻击行为组成的特征集合远离正常特征集合，通过不同的特征类型区分攻击行为和正常行为[14]。划分通信网络中的若干组行为为 i 个类型，各组类型中的特征相似度较高，而不同类特征之间的相似度较低。随机将 i 个类型作为初始中心，分别计算剩余类与设定中心的距离，将距离较近的类规划为一个族群，并重新计算初始中心，在不断地收敛过程中以均方误差作为标准函数，获取最佳的中心搜索位置：

$$mse=\sum_{i=1}^{\sqrt{t}}\sum_{u\in y_i}\|u-u_i\|^2\times\varepsilon\times k_{ls}(a)$$

$$mse'=\sum_{i=1}^{\sqrt{t}}\sum_{u\in y_i}\|u-u_i\|^2\times\varphi\times h_{ls}(a) \tag{5}$$

公式中：标准函数定义为 mse[15]，族群表示为 y_i，u_i 表示 y_i 的中心，u 为划分空间中的某一个对象，ε 为标准函数位置转换系数，φ 为标准函数速度转换系数，i 为恶意攻击代码类型，对 i 个类型数量进行组合编码，目前 i 的最大值为 $\sqrt{t}$，t 表示为样本的总数，因此 i 的取值范围为 $[2,\sqrt{t}]$。通过标准函数的设定对攻击代码粒子的攻击速度和粒子位置判断，以此评价恶意攻击代码粒子的恶意攻击行为[16]。当标准函数 mse 的结果越小表明攻击效果越明显，则恶意行为特征分类的程度越高，以粒子群优化算法中适应度函数定义攻击粒子的最优解，如下：

$$fit=\frac{1}{mse+mse'} \tag{6}$$

公式中：fit 为粒子群优化算法适用度函数[17]。当 mse，mse' 值越小时表示攻击代码粒子的适用度值越大，则其寻找到的粒子位置就越容易攻击正常节点，即可将其确定为

攻击代码的优先选择的位置，否则当适用度值越小，则表示该粒子位置比攻击代码经过的所有位置都不适合攻击，即该粒子当前位置为安全位置，代码重新选择位置进行攻击，因此，可以通过计算恶意攻击代码的适用度值来确定其攻击的最佳粒子位置，获得最优解。具体最优解寻找过程如图 2 所示。

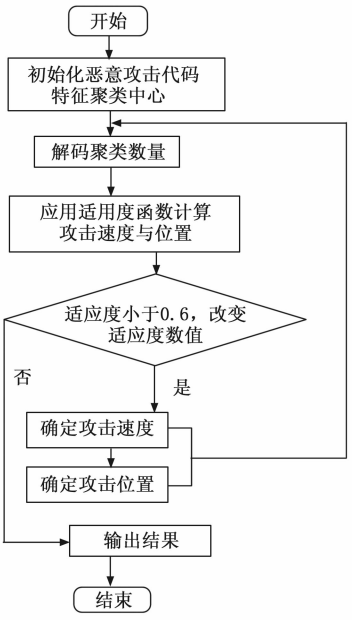

图 2　粒子群优化算法的最优解寻找过程示意图

根据图 2 内容所示，将粒子群优化算法与攻击原理相结合，通过该算法中的适用度函数能够判断恶意攻击代码的粒子个体极值与全局极值，并且随着适用度函数的不断变化，粒子的两个极值会进行更新，最后得到的粒子最优解即为恶意代码确定的攻击位置[18]。以恶意攻击代码的适应度数值判断攻击粒子的位置，能够有效区分正常行为代码粒子位置和恶意攻击代码粒子位置，

$$h'_{ls}(a)=h_{ls}(a),fit\geqslant\psi \tag{7}$$

$$k'_{ls}(a)=k_{ls}(a),fit\geqslant\psi \tag{8}$$

以公式（7）和公式（8）获取的恶意攻击代码粒子的攻击速度和粒子位置为基础，构建恶意攻击代码特征集，即最终的最优解，公式为：

$$B=\{h'_{ls}(a),k'_{ls}(a)\} \tag{9}$$

至此实现恶意攻击代码更新最优解获取，为检测恶意攻击代码奠定基础。

4　基于 PSO-KM 聚类分析检测恶意攻击代码

上述完成了 PSO 算法寻找恶意攻击代码更新最优解，但是此时判断出的恶意攻击代码粒子位置和速度仍存在一定的误差，直接应用的话，无法达到最佳效果，因此，为了进一步提高恶意攻击代码检测的准确性，在上述 PSO 算法寻找恶意攻击代码更新最优解的基础上，引入 KM 聚类分析，检测恶意攻击代码。在检测过程中，外在网络行为特征提取是检测重点，将上文划分的网络通信协议层存在的攻击代码按照 4 种形式进行划分，划分后的恶意攻击代码可以表示为活跃代码、故障代码、扫描代码以及页面代码。通过 4 种网络代码特征对恶意代码进行聚类分析，将其作为检测前提记录原始数据层的网络数据流[19]。

由于网络数据流具有粒度特性，在记录和存储通信数据时，其操作包含传输时间、IP 源地址、IP 目的地址、IP 端口和故障信号这 5 个属性。不同数据处理过程中均可以确定 IP 地址的分类情况，将局域网的 IP 地址定义为审计地址，对可疑的外网位置定义为检测地址。以连接信号确定地址的所属关系类型，如下：

$$\begin{cases}wanIPfail\in IP\\ wanIPact\in IP\\ wanIPfail\cap wanIPact\cap wanIP\in\varphi\end{cases} \tag{10}$$

式中，φ 表示孤立判断阈值。$wanIPfail$ 表示不活跃的外网 IP 地址[20]。$wanIPact$ 表示活跃的外网 IP 地址。$wanIP$ 表示所有可疑的外网 IP 地址。根据网络行为特性类型设定特征提取模块，分别为数据获取模块和网络行为特征提取模块，对 IP 地址是否为攻击源头进行判断。

在获取模块内采集通信网络中的数据流量，并将网络流量数据传输到行为特征模块之中，对预先分类的 4 种网络行为进行特征提取，若网络行为特征值较低则说明其来自不活跃的外网，若其行为特征值较高则属于活跃外网，一般会将活跃度较高的特征值判断为异常行为进行检测。输入 IP 实际特征集合为 $Q=[Q_E]_{E=1}^{W}$，特征聚类中心集合为 $R=\{R_1,R_2,\ldots,R_T\}$，对特征数据进行排序：

$$Y=\frac{W+1}{2} \tag{11}$$

$$Y=\frac{W}{2} \tag{12}$$

式中，Y 表示特征集合中的数据的中位数[21]。上式（8）表示 W 为奇数时的排序方式，否则通过公式（9）计算。在聚类中心的排序为 $|R_1|<|R_2|<|R_3|<,\ldots,<|R_T|$ 情况下，设定恶意代码惯性系数的动态权重，平衡恶意代码的全局检测：

$$\delta=\delta_{\max}-\frac{\delta_{\max}-\delta_{\min}}{U_{\max}}*U \tag{13}$$

式中，惯性系数的动态权重上下限分别为 $\delta_{\max}$ 和 $\delta_{\min}$。U 表示孤立点样本。$U_{\max}$ 表示样本最大值。基于此对 IP 地址中的恶意代码网络行为特征进行跟踪，结合 PSO 算法寻找恶意攻击代码更新最优解，实现恶意代码最终位置 I 的检测：

$$I=\frac{\sum_{T=1}^{A}\left\{\frac{\sum_{G=1}^{F_T}\delta DIS(R_T,H_{TG})}{F_T}\right\}}{A}+\gamma\frac{\sum_{Z=1}^{C}\left\{\frac{\sum_{G=1}^{F_Z}\delta DIS(R_Z,B)}{F_Z}\right\}}{C} \tag{14}$$

式中，H_{TG} 表示第 T 条流量中第 G 个网络行为的特征集合。R_T 为第 T 个聚类中心。$DIS(R_T,H_{TG})$ 表示数据矢量 H_{TG} 和 R_T 之间的距离。F_T 为单个数据集合数目。A 为集群的数目。

γ 为转换系数。F_C 为单个恶意攻击代码数据集合数目。R_C 为第 C 个聚类中心。根据聚类结果区分恶意攻击代码和正常数据，即检测出恶意攻击代码。至此，基于 PSO-KM 聚类分析方法实现通信网络恶意攻击代码检测方法设计。该方法为了确保检测数据准确性，确定了通信网络流动轨迹中恶意攻击行为特征，并且归一化处理通信网络恶意攻击行为特征，在此基础上，引入了 PSO 算法，通过该算法寻找恶意攻击代码更新最优解，为了提高检测准确性和效率，再次引入了 KM 聚类分析算法，通过 KM 聚类优化 PSO 算法，构建了 PSO-KM 聚类分析方法的恶意攻击代码检测方法。该方法具备了多个算法的优点，有效提高了检测准确性，降低检测误报率。

5 实验测试分析

5.1 实验方案

为了验证设计方法的有效性和应用性能，设计对比分析实验。考虑实验的有效性，设计实验方案，具体方案如下所示：

1）选择恶意攻击数据。明确研究对象，即恶意代码攻击样本集，并且给出训练集合和测试集合分类信息；

2）实验参数设置。实验参数的不同有可能影响实验结果，因此，为了避免实验参数的影响，提前设置设计方法的实验参数具体数值；

3）实验性能指标。验证设计方法的性能，需要具体的实验性能指标，通过具体指标反映方法的性能，该实验以恶意代码检测个数统计结果和检测效果为性能指标，其中，检测效果分为识别率和误报率，并且给出具体计算公式；

4）分析恶意代码检测结果。通过上述指标，以择基于遗传算法的检测方法、基于灰狼算法的检测方法以及基于规则库的检测方法作为对照组，与本文方法进行对比分析。

按照上述设计的实验方案开展实验。

5.2 选择恶意攻击数据

上文中通过 PSO-KM 聚类分析方法设计了一个新的检测方法，为验证该方法能够完成通信网络恶意攻击代码的有效检测，采用对比测试方法进行论证。分别选择基于遗传算法的检测方法、基于灰狼算法的检测方法以及基于规则库的检测方法作为对照组，分别与本文方法进行比较，在不同类型的数据包样本中完成测试。

为保证测试的准确性和真实性，以通信网络中实时采集的数据作为测试样本，选择 20 000 组正常数据作为请求集合，43 008 组数据作为攻击样本请求集合，对 43 008 条攻击样本训练集中的代码进行统计，恶意攻击代码具体情况如表 2 所示。

根据表 2 内容所示，不同类型恶意攻击代码的数量有所不同，对恶意攻击样本集合中的攻击代码进行分词，提取关键的攻击代码构建词组集合，并将其放置在 ACUNETIX-VULNERABILITY-SCANNER 漏洞扫描器进行跟踪，建立一个纯度较高的训练样本集合。通过 SKLEARN.

表 2 恶意代码攻击样本集合

序号	恶意攻击代码名称	数量/组
1	We	453
2	=sdf	2 089
3	Jaerf	2 520
4	=67kj	2 631
5	Gha	2 214
6	%3^g	1 234
7	<rgk	1 929
8	Ga46&.	3 102
9	=<hi	2 123
10	Ih19^	2 079
11	Hgzcb64^	2 508
12	v/l^	2 793
13	v/g^	2 705
14	thgs	2 096
15	.aa	2 107
16	.z	1 972
17	IeSearchBar	2 295
18	Gpigeon	1 037
19	LoveLetter	2 650
20	Fujacks	2 471

CROSS-VALIDATION 模块随机选择 10 000 组样本分类样本集合（14 类恶意攻击代码），训练集合和测试集合的随机分类情况如下：

1）A1 分类：70%的训练样本集合，30%的测试样本集合。

2）A2 分类：90%的训练样本集合，10%的测试样本集合。

3）A3 分类：10%的训练样本集合，90%的测试样本集合。

4）A4 分类：30%的训练样本集合，70%的测试样本集合。

将上述分类完毕的样本集合导入至 Matlab 测试平台中，分别连接选择的四组检测方法，按照两个阶段完成测试：第一阶段验证不同方法的识别量，即每组检测方法在对恶意代码攻击检测时识别出的具体数量，一般情况下认定能够在较少识别个数下完成恶意攻击代码识别，则表明该检测方法的检测效率高。并统计检测方法未识别出的恶意代码数量，其未识别的恶意代码数量越少，则说明检测方法更加有效。第二阶段验证不同方法的识别率和误报率，即对所有识别出来的代码进行对照，验证在所有识别出来的代码中，是否为真正的攻击代码，当识别率越高说明检测方法的准确率就越高，误报率越低也能够说明检测方法的准确性越好。按照不同的测试阶段完成检测，验证不同方法的有效性。

5.3 实验参数设置

实验参数的设置在实验中占据重要位置，这是因为实验参数数据可能影响实验结果，导致实验分析不准确，因此，

为了避免影响，设置准确的实验参数，具体如表 3 所示。

表 3　实验参数

序号	参数	数值
1	v	0.679
2	c	0.537
3	p	43 286
4	a	(0,1)
5	i	20
6	φ	1.254
7	A	20
8	γ	0.713
9	T	20

按照上述数据设置实验过程的参数。

5.4　实验性能指标

实验过程选择恶意代码检测个数统计结果和检测效果作为具体的实验性能指标，其中恶意代码检测个数统计结果按照不同的分类情况对各组检测方法下识别的恶意代码个数和未识别的代码个数进行统计，对比分析不同方法的识别个数与未识别个数，该指标通过计算机自带软禁直接统计。检测效果分为识别率和误报率，计算公式为：

1）识别率：

$$q = \frac{w}{e} \times 100\% \tag{15}$$

式中，q 表示识别率；w 表示正确识别的代码数量；e 为恶意攻击代码的总计数量。

2）误报率：

$$r = \frac{t}{e} \times 100\% \tag{16}$$

式中，r 表示误报率；t 表示将正常数据判断为恶意攻击代码的数量。

上述实验指标中，恶意代码检测个数统计结果的恶意代码个数和检测效果的识别率越高，则说明检测方法的性能越好，恶意代码检测个数统计结果的未识别个数和检测效果的误报率越低，则说明检测方法的应用效果越佳。

5.5　分析恶意代码检测结果

根据上文中设定的内容，按照不同的分类情况对各组检测方法下识别的恶意代码个数和未识别的代码个数进行统计，如图 3 所示。

根据表中内容所示应用不同的检测方法后，对恶意代码的识别结果各不相同，其中本文方法只需要较少的时间就可以完成较多恶意攻击代码的识别，在 5 中分类中，基本在 30 s 内完成恶意攻击代码检查，并且识别的总恶意攻击代码数量最多，其中 A2 分类下，恶意攻击代码检测数量基本达到了 1 000 个，其他分类下，检测出来的数量与实际数量基本一致。而对比另外 3 种方法可知，其他方法均与实际需要识别的恶意攻击代码数量差距较大，仅本文

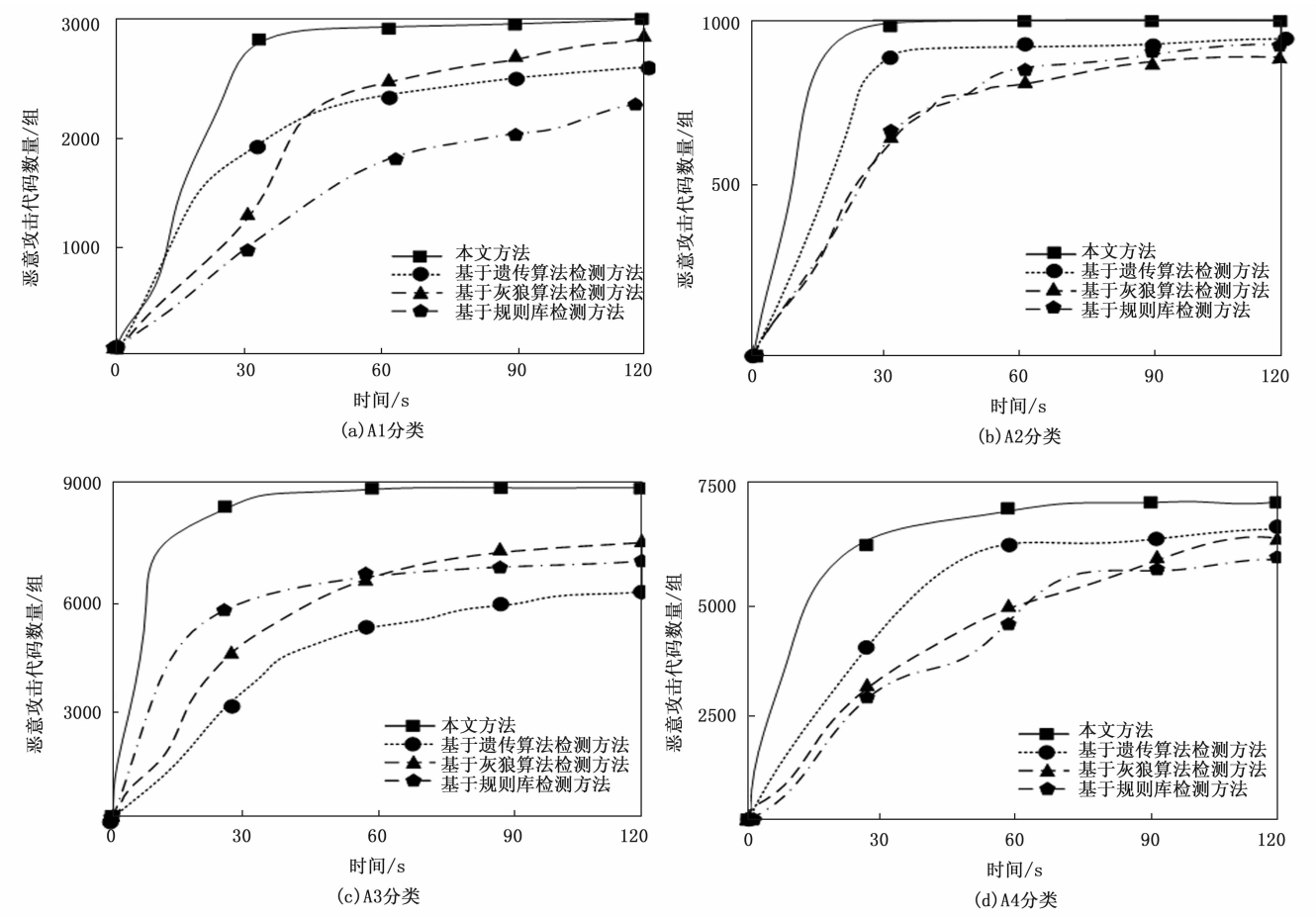

图 3　恶意代码检测个数统计结果

方法的识别数量与实际数量趋近，甚至全部识别出所有恶意攻击代码，3 种方法的识别时间大概在 60 s 左右，60 s 以后识别出的恶意攻击代码数量基本不在增加。综合比较下，本文方法的效率更佳，并且识别恶意攻击代码的数量最多。

在此基础上，分析各组检测方法的恶意代码识别率和误报率。以图 3 中给出的识别个数计算各组方法的识别率和误报率，结果如图 4 所示。

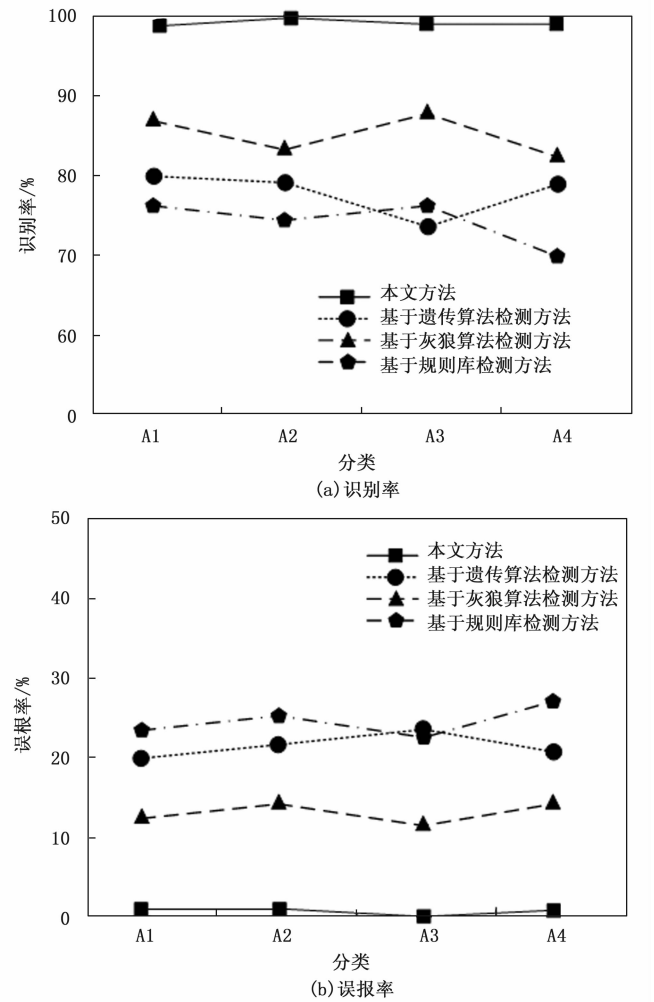

图 4 不同方法下检测效果

根据图 4 内容所示，仅本文方法的识别率在 95.0%以上，且识别率最高值接近 99.7%，而检测的误报率可以控制在 0.4%之内，并且综合图 3 中的识别数量，说明本文方法更具有泛化能力，能够在较少的识别数据量中完成恶意代码的检测，具有较高精确度。基于规则库检测方法和基于遗传算法检测方法包含有非攻击代码，两者的误报率较大，均在 20%以上，并且最高识别率仅为 80.0%，基于灰狼算法检测方法由于识别的个数较多，在进行样本训练和获取中概率值范围变宽，则误报率也较低，但与本文方法相比仍存在一定差距，其识别率最高仅为 88.2%，并且误报率最低为 12.7%。对比 4 种方法，本文方法的识别率提高了 6.0%以上，并且误报率降低了 10.0%以上，由此可知，本文方法有效提高了恶意攻击代码的识率，降低了误报率，从而提高了应用性能。

综合结果可知：本文方法能够对恶意攻击代码进行特征分析，将同属于某个类型的恶意攻击代码进行识别和分类。在对安全测试中的恶意攻击代码进行关联分析后获取特征集合，加强了恶意代码的特征描述，提高了恶意攻击行为检测的识别率，降低了恶意攻击行为代码的误报率，从而保证了检测效果，具有应用价值，应用价值主要体现在电力通信网络、航海导航、广播电视等领域。

6 结束语

为了提高通信网络的恶意代码检测效果，并且考虑 PSO-KM 聚类分析的优异性和适用性，本文引入了 PSO-KM 聚类分析，设计了一种新的通信网络恶意攻击代码检测方法。该方法在引入 PSO-KM 聚类分析前，为了提高检测准确性，确定了通信网络流动轨迹中恶意攻击行为特征，归一化处理通信网络恶意行为特征，并且通过适应度函数寻找攻击代码更新最优解，结合 PSO-KM 聚类分析实现恶意代码检测。同时，在实验论证的基础上检验了新方法应用的效果，具有高度准确性和较低的误报率，其中识别率达到了 95.0%以上，误报率最高值仅为 0.4%，与对比方法相比，本文方法的识别率提高了 6.0%以上，误报率降低了 10.0%以上，该方法有效提高了识别率和降低了误报率。但由于本文研究时间有限，在测试过程中仍存在少许不足之处，主要是恶意代码的特征数量选取有限，对比测试中的数据准备不够充足，后续研究中会设定更多的恶意代码，为检测方法的全面性应用提供理论支持。

参考文献：

[1] 杨 静，郭韦昱，杨文彬. 基于深度森林的可见光通信网络恶意代码识别研究 [J]. 激光杂志，2022，43 (10)：150－154.

[2] 张 杰，景 雯，陈 富. 基于被动分簇算法的即时通信网络协议漏洞检测 [J]. 吉林大学学报 (工学版)，2021，51 (6)：2253－2258.

[3] 杨望，高明哲，蒋 婷. 一种基于多特征集成学习的恶意代码静态检测框架 [J]. 计算机研究与发展，2021，58 (5)：1021－1034.

[4] 吕 靖，齐海迪，李宝德. 基于扩展置信规则库的海盗袭击事件风险预测 [J]. 交通运输系统工程与信息，2022，22 (3)：247－254.

[5] 张大明，徐嘉庆，赵彦清，等. 基于停滞检测的双向搜索灰狼优化算法 [J]. 计算机应用研究，2022，39 (6)：1725－1730.

[6] 张英贵，陆 强，高 全，等. 基于关键站点识别的区域轨道交通路网抗干扰性研究 [J]. 铁道科学与工程学报，2022，19 (7)：1845－1853.

[7] 王 瑛，张 泽，张 滢，等. 基于多属性决策的 PCAS-SNIF 关键节点分析 [J]. 空军工程大学学报 (自然科学版)，2021，22 (6)：90－96.

[8] 曹鹏宇，杨承志，石礼盟，等. 基于 PSO-DBSCAN 和 SC-

GAN 的未知雷达信号处理方法 [J]. 系统工程与电子技术，2022，44 (4)：1158－1165.
[9] 李　清. 基于改进 PSO-PFCM 聚类算法的电力大数据异常检测方法 [J]. 电力系统保护与控制，2021，49 (18)：161－166.
[10] 刘　云，肖　添. IP 黑名单关联聚类算法对恶意簇检测的优化研究 [J]. 四川大学学报 (自然科学版)，2021，58 (1)：97－104.
[11] 王小英，刘庆杰，庞国莉. 恶意代码攻击下多业务通信网络安全响应仿真 [J]. 计算机仿真，2020，37 (10)：137－141.
[12] 张代华，沈　勇，章翔飞，等. 基于数据挖掘的网络入侵数据自主防御仿真 [J]. 计算机仿真，2020，37 (10)：263－267.
[13] 陈　霖，许爱东，蒋屹新，等. 基于动态增量聚类分析的电力信息网络攻击模式识别算法 [J]. 南方电网技术，2020，14 (8)：25－32.
[14] 庞传军，余建明，冯长有，等. 基于 LSTM 自动编码器的电力负荷聚类建模及特性分析 [J]. 电力系统自动化，2020，44 (23)：57－63.
[15] 夏景明，丁春健，谈　玲. 基于灰狼算法的深度信念网络入侵检测方法 [J]. 计算机工程与设计，2020，41 (6)：1534－1539.
[16] 韩　超，段纬然，贾长治，等. 基于改进多种群 PSO 算法的火炮随动系统调节器参数优化 [J]. 火力与指挥控制，2020，45 (6)：56－61.
[17] 牛在森，伊华伟，李晓会，等. 融合层次聚类和粒子群优化的鲁棒推荐算法 [J]. 小型微型计算机系统，2020，41 (1)：66－71.
[18] 何　俊，张云飞，张德海. 基于 Petri 网的数据清洗规则链自动组合与检测 [J]. 计算机工程，2020，46 (11)：124－131.
[19] 解　滨，董新玉，梁皓伟. 基于三支动态阈值 K-means 聚类的入侵检测算法 [J]. 郑州大学学报 (理学版)，2020，52 (2)：64－70.
[20] 朱　迪，陈丹伟. 基于密度聚类和随机森林的移动应用识别技术 [J]. 计算机工程与应用，2020，56 (4)：63－68.
[21] 胡艳香，孙　颖，张雪英，等. 基于聚类 PSO-LSSVM 模型的 PAD 维度预测 [J]. 计算机应用研究，2020，37 (4)：994－998.

(上接第 7 页)

1) 通过数据采集终端和串口服务器，实现了涉电终端设备，无差别接入监测系统，并实现数字化。

2) 通过融合设计 C/S 及 B/S 架构，实现了数据从采集到存储，再到处理和显示的整个流程，并自行设计了数据采集终端和上位机服务器之间的网络通信协议。上位机服务器通过 websocket 实现界面实时刷新。

3) 提出了将数据回归分析应用于数据处理分析的方法。

4) 满足了航天测控站对供电系统监测的需求。

解决了航天测控站电力系统有价值数据挖掘不足、缺乏电力系统数据集中显示手段的问题，并提出了电力系统数据潜在的智能化应用方向，为确保电力系统的稳定可靠运行提出了可行方案，为航天测控任务的正常实施提供了有力支撑。

参考文献：

[1] 张晓兰，谷永刚，张少强，等. 智能监测系统下的线性电网状态估计研究 [J]. 电测与仪表，2021，58 (10)：112－117.
[2] 徐一晨，杨吉军，詹发军，等. 电压骤降故障下电网线损智能监测系统设计 [J]. 自动化与仪器仪表，2019 (12)：57－60.
[3] 刘　琨，黄明辉，李一泉，等. 智能变电站故障信息模型与继电保护在线监测方法 [J]. 电力自动化设备，2019，38 (2)：210－216.
[4] 刘　铸. 燃煤热电厂智能化适应性供热监测监控系统开发与应用 [D]. 太原：太原理工大学，2022.
[5] 翁慧颖. 电涌保护器及其智能监测系统研究 [D]. 杭州：浙江大学，2011.
[6] 何健波. 一种配电柜电参数及供电环境智能监测系统 [P]. 中国：201621093230.8，2017.
[7] 徐建军，等. 黄陵矿区供电系统智能化建设技术与实践 [J]. 智能矿山，2022，3 (10)：23－33.
[8] 刘　波，等. 井下供电监控系统的设计 [J]. 能源技术与管理，2022，47 (5)：131－133.
[9] 范学刚，等. 矿井地面供电系统改造设计与应用 [J]. 机械研究与应用，2022，35 (2)：129－132.
[10] 李明生，等. 矿井智能供电监控系统设计及应用研究 [J]. 冶金管理，2023 (3)：61－63.
[11] 穆艳祥，等. 煤矿供电系统的智能化建设研究 [J]. 内蒙古煤炭经济，2022 (6)：108－110.
[12] 杨记根，等. 智能化供电技术在煤矿供电系统中的应用 [J]. 智能矿山，2022 (3)：23－33.
[13] 石　岩，等. 多元传感器智能监测系统在高铁监测中的应用 [J]. 地理空间信息，2022，20 (11)：107－110.
[14] 代小红，等. 基于无线传感器的智能监测输液系统设计 [J]. 压电与声光，2014，36 (6)：972－976.
[15] 温从众，等. 智能电网监测传感器设计与研究 [J]. 传感技术学报，2023，36 (2)：210－217.
[16] 李贤良，等. 光纤光栅传感器在变压器多参量智能监测中的应用研究 [J]. 电力系统保护与控制，2023，51 (6)：154－160.
[17] 余　刃，等. 一种用于旋转设备运行状态无线监测的智能无线振动传感器设计 [J]. 核动力工程，2020，41 (3)：221－226.
[18] 刘　浩，等. 基于传感器技术的智能电网输电线路监测系统 [J]. 电子设计工程，2023，31 (10)：172－176.
[19] 徐俊辉，等. 智能无线传感器在振动监测系统中的实际应用分析 [J]. 石油石化物资采购，2023 (3)：181－183.
[20] 刘　宇，等. 面向配电室智能监测的传感器巡检机器人设计 [J]. 计算技术与自动化，2020，39 (2)：67－72.