彭懋磊,刘可,李垠,等.基于互联网的震后虚假信息入侵实时检测系统设计[J].地震工程学报,2020,42(1):270-275.doi:10.3969/j.issn.1000-0844.2020.01.270

PENG Maolei,LIU Ke,LI Yin,et al.Design of a Real-Time Detection System for False Post-Earthquake Information Intrusion Based on the Internet[J].China Earthquake Engineering Journal,2020,42(1):270-275.doi:10.3969/j.issn.1000-0844.2020.01.270

# 基于互联网的震后虚假信息入侵实时检测系统设计

彭懋磊,刘 可,李 垠,吕 筱

(中国地震局地震研究所(地震大地测量重点实验室),湖北 武汉 430071)

**摘要**:虚假震后信息入侵严重干扰抗震救灾与灾后重建进程,为防止虚假震后信息在互联网中大肆传播,需对震后虚假信息进行精准识别,设计基于互联网的震后虚假信息入侵实时检测系统。互联网以 Web 服务器为载体传播信息,数据包捕获模块采用 WinPcap 技术,并从中获取有效数据包,在协议解析模块中完成数据包 TCP/IP 协议、CMP 协议、UDP 协议解码工作,解码后的数据输入到基于 SVM 的震后虚假信息检测模型中。SVM 模型寻求最优超平面将震后信息分为虚假与真实两个类别,根据该结果完成震后虚假信息入侵实时检测。由实验结果可知该系统检测震后虚假信息入侵误报率低于 3%,相比同类型系统具有接收数据能力强、效率高的优势,对精准检测震后虚假信息具有重要意义。

**关键词**:互联网;震后;虚假信息;入侵;协议解析;实时检测

**中图分类号**:TP393.08 **文献标志码**:A **文章编号**:1000-0844(2020)01-0270-06

DOI:10.3969/j.issn.1000-0844.2020.01.270

# Design of a Real-Time Detection System for False Post-Earthquake Information Intrusion Based on the Internet

PENG Maolei, LIU Ke, LI Yin, LÜ Xiao

(*Key Laboratory of Earthquake Geodesy, Institute of Seismology, China Earthquake Administration, Wuhan 430071, Hubei, China*)

**Abstract**: False post-earthquake information intrusion seriously interferes with earthquake relief and postdisaster reconstruction. To prevent the spread of false post-earthquake information on the Internet, this study designs a real-time detection system. Web servers are used as carriers to transmit information on the Internet. Thus, the WinPcap technology is used herein to capture valid data packets. The transmission control protocol/internet protocol, control message protocol, and user datagram protocol are decoded in the protocol analysis module. The decoded data

**收稿日期**:2019-04-23

**基金项目**:中国地震局监测、预测、科研三结合项目(CEA-JC/3JH-161701);中国地震局地震研究所基本科研业务费专项资助项目和中国地震局地壳应力研究所基本科研业务费专项资助项目(IS201426146)

**第一作者简介**:彭懋磊(1981-),男,湖北荆州人,硕士,工程师,主要研究方向:地震信息网络架构、服务、安全。E-mail:pengmaolei@163.com。

are used inputs in the false post-earthquake information detection model based on support vector machine. The model seeks the optimal hyperplane to divide the post-earthquake information into false and true. Results indicate the completion of the real-time detection of false post-earthquake information intrusion. The false alarm rate of this system is less than 3%. Different from related systems, the proposed system can receive data with high efficiency and is of great importance in accurately detecting false post-earthquake information.

**Keywords**: Internet; post-earthquake; false information; intrusion; protocol analysis; real-time detection

## 0　引言

发生地震灾害后,震后灾情信息在互联网中迅速传播蔓延,其中掺杂大量虚假信息,影响灾民精神状态、阻碍灾后快速重建[1]。因此,地震灾害发生后,相关部门应采取措施防止虚假震后信息在互联网中蔓延[2]。现有震后虚假信息入侵实时检测系统较多,例如文献[3]提出了结合特征分析和 Svm 优化的 Web 入侵检测系统,结合 http 协议分析,选取密切相关的数据特征,利用现阶段成熟稳健的 SVM 算法进行学习分类,使用网格搜索法进行参数优化,同时使用序列最小最优化 SMO 算法加速运算,从而有效地提高了入侵检测的效率和准确率。由于计算量较大,导致系统时间开销与能量消耗大,检测性能较低;文献[4]提出了基于神经网络的入侵检测系统,在传统 BP 算法基础上,采用自动变速率学习法,引入遗忘因子、随机优化算子,并将其用于网络入侵检测系统。这种系统检测效率较高,但常因为特征空间维度高问题导致检测精度低,不能作为高质量的虚假信息检测系统使用。文献[5]针对异常序列检测提出一种非法入侵识别算法,其识别异常信息能力强,但仅限于非法入侵信息的识别,不能准确判断信息的真实与虚假性质,不符合震后虚假信息检测标准。针对上述检测方法与系统存在的缺点,本文基于互联网设计新的震后虚假信息入侵实时检测系统,解决了上述检测方法时间开销大、检测精度低的问题。

## 1　震后虚假信息入侵实时检测系统设计

### 1.1　系统总体结构

图 1 为基于互联网的震后虚假信息入侵实时检测系统的总体结构。由图 1 能够看出,系统主要由 Web 服务器、数据包捕获模块、协议解析模块、数据包控制台等部分构成[6]。系统各部分功能如下:Web 服务器是互联网传播震后信息的载体,Web 服务器信息、相关模块的日志信息存储在数据包控制台中,数据包控制台控制各环节数据包发送量,信息安全管理员可查看图形化的日志信息,为有效分析虚假信息提供依据;数据包捕获模块负责采集入侵端发送的震后虚假信息[7];协议解析模块对捕获的数据包进行协议解码,解码后的数据接口可供检测系统使用;SVM 分类模型将数据包信息作为输入量,分类结果作为输出量,据此检测震后虚假信息入侵情况。

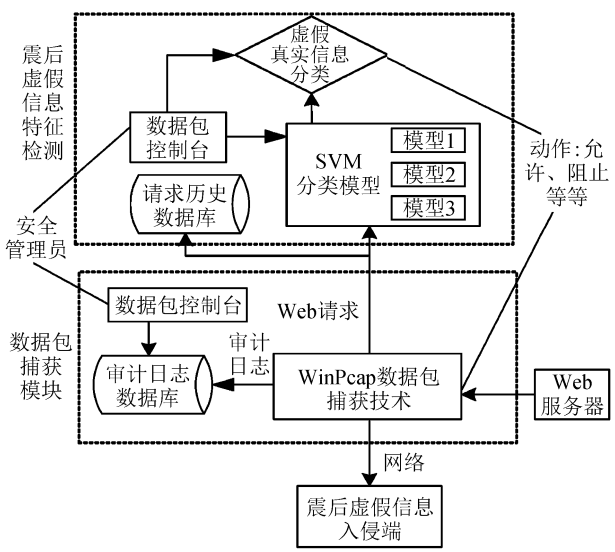


图 1　基于互联网的震后虚假信息入侵实时检测系统总体结构

Fig.1　The overall structure of a real-time detection system for post-earthquake false information intrusion based on Internet

### 1.2　系统模块设计

(1) Web 服务器

互联网由多个 Web 服务器构成,服务器相互连接形成内容丰富的互联网。互联网通过 Web 服务器实现震后虚假信息入侵检测[8],系统采用的 Web 服务器结构如图 2 所示。

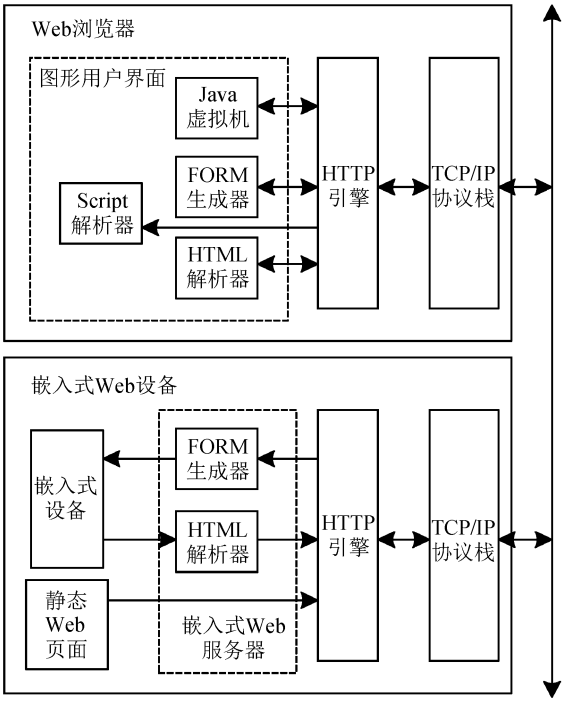


图 2 Web 服务器结构

Fig.2 Web server architecture

由图 2 能够看出，Web 服务器包括 Web 设备与 Web 浏览器两部分；TCP/IP 协议、HTTP 协议是连接 Web 设备与 Web 浏览器的纽带，前者负责底层通信，后者负责顶层通信；Web 浏览器采用 Script 解析器、HTML 解析器完成浏览器脚本与 HTML 语言的解析工作；Web 服务器接收 Form 生成器生成的动态 Form，接着由 Form 解析器解析动态 Form，与 Web 设备的应用管理程序通道协作配置，控制 Web 设备。基于互联网检测震后虚假信息时，Web 设备与 Web 服务器应用程序的通信，体现在 Web 浏览器端生成的 HTML 文档。

(2) 数据包捕获模块

在互联网中获取震后信息是数据包捕获模块的主要任务，数据包捕获模块是系统的基本处理模块[9]。模块功能如下[10]：从互联网中采集震后信息数据包，提取受保护网络的数据包供协议分析模块进行分析，是系统的数据来源。网络中信息流不断增加，以此数据包捕获模块必须具备稳定、可靠的性能，降低系统的误报率，提升系统检测率[11]。数据包捕获模块使用 WinPcap 数据包捕获技术，因而具备优秀的数据包捕获机制，信息过滤技术可自动过滤用户不关注的数据包，大大提升系统获取网络数据包的性能。数据包捕获模块结构见图 3，由图 3 可知，该模块主要由低级动态链接库、高级动态链接库、内核级网络组包过滤器构成。

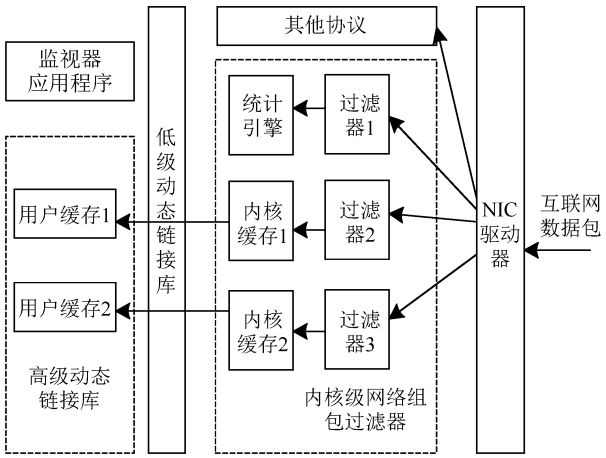


图 3 数据包捕获模块结构

Fig.3 Structure of packet capture module

(3) 协议解析模块

归纳系统协议解析模块工作流程如图 4 所示。由图 4 可知，数据包捕获模块获取数据包后，经过以太网将数据包发送至协议解析模块，进行 TCP/IP 协议、CMP 协议、UDP 协议解码[12]，解码后的数据包接口符合检测系统要求，以此为基础进行震后虚假信息入侵检测。

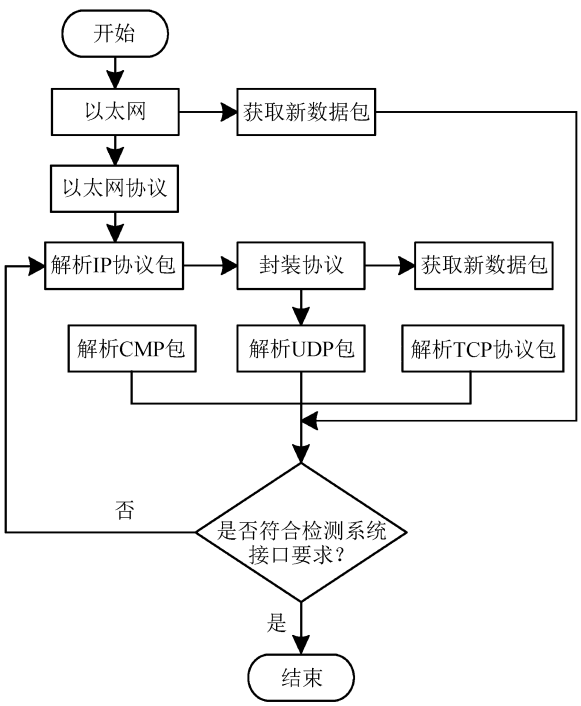


图 4 协议解析模块工作流程

Fig.4 Workflow of protocol analysis module

### 1.3 基于 SVM 的震后虚假信息检测模型

(1) 灾害信息集合

假设信息类别集合为 $G=\{g_1,g_2,g_3,\cdots,g_n\}$，

其中,$g_i$ 是震后灾情信息,包括地震发生的时间、灾害程度、灾害范围等内容。定义信息发布用户的信誉值为 $H_e$,是可信度的体现。用户的 $H_e$ 为可变化的,根据用户历史消息发布程度进行判定。$H_e$ 值越小、可信度越差,但是信用程度与信息真实程度无线性关系。

震后发布的灾害信息集合为 $\Phi=\{0,1\}$,"0" 表示信息属实,"1" 表示信息虚假。检测震后虚假信息首先应提取信息特征,信息特征存在多维度性质,定义 $X=\{x_1,x_2,x_3,\cdots,x_n\}$ 为特征向量集合,是震后信息、发布者信息的关键体现。$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$ 是与之对应的数据集合,$y_i\in\Phi$ 为针对样本 $x_i$ 的真实结果。

(2) 虚假震后信息入侵检测

采用 SVM 模型将入侵的震后信息划分成虚假信息、真实信息两个类别。检测虚假震后信息入侵过程中,全部原始信息视为高维空间内的圆点,用一个最优平面最大间隔划分数据点[13],过程见图 5。

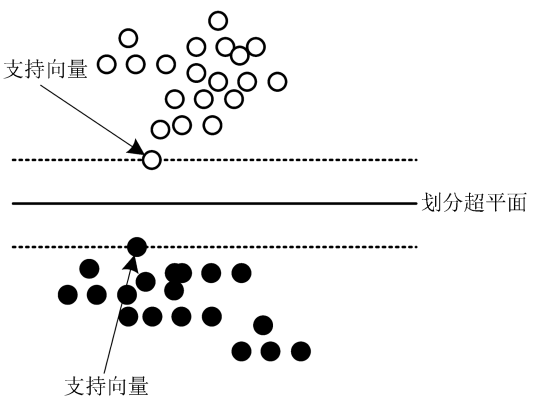


图 5　SVM 超平面划分

Fig.5　SVM hyperplane partition

归纳基于 SVM 模型检测震后虚假信息过程如下:

首先,提取震后信息的多维特征向量[14],用 $x$ 表示。地震灾害的程度 $e$、发生时间 $s$、受灾人数 $d$、经济损失量 $a$、消息发布用户信誉度 $H_e$ 等都属于特征向量的维度。

那么虚假消息的特征向量如下:

$$x=[e;s;d;a;H_e;l;t_0;c] \tag{1}$$

其中,$SVM$ 模型训练的输入量为多维特征向量 $x$。经过选取核函数、优化参数、较差验证等步骤获取震后虚假信息检测的最优超平面[15],如公式(2) 所示:

$$\begin{aligned}&\min_{w,b,\zeta_i}\frac{1}{2}\|w\|^2+D\sum_{i=1}^{m}\zeta_i\\&\text{s,t. } y_i(w^Tx_i+u)\geqslant 1-\zeta_i\\&\zeta_i\geqslant 0,i=1,2,\cdots,m.\end{aligned} \tag{2}$$

式中:$D$ 是惩罚因子;$w$ 为特征向量权重;$xi$ 为优化后的多维特征向量;$\zeta_i$ 为特征检测的松弛变量;$u$ 为偏置量。

采用高斯核函数优化SVM模型的维数,确保精准划分虚假信息与真实信息,$\sigma$ 为方差,核函数形式如下:

$$K(x,x_i)=\exp(-\|x,x_i\|^2/2\sigma^2) \tag{3}$$

根据上述构建的 SVM 分类模型判别互联网中震后信息是否为虚假信息,实现震后虚假信息入侵检测。

## 2　实验分析

为验证本文系统检测震后虚假信息的有效性,展开实验研究。相关部门提供某次地震的真实数据信息,将数据信息归纳成 5 个数据包,分别命名为 A、B、C、D、E,各包含 500 条数据信息。人为制造虚假震后信息,使用 Hgod 入侵软件根据实验需要进行实时入侵。选取 10 台服务器组成局域网;控制端共需两台计算机,配置如下:Windows XP 系统,500 G 存储容量。为突出实验效果,采用 Web 入侵检测系统、基于神经网络的入侵检测系统进行对比实验。

### 2.1　检测率分析

实验进行中,人为制造虚假信息进行网络入侵,本文系统检测震后虚假信息入侵性能如表 1 所列。

**表 1　本文系统检测震后虚假信息入侵性能**

**Table 1　Intrusion performance of false post-earthquake information detected by the proposed system**

| 数据包名称 | 检测率/% | 系统开销(时间)/ms | 系统开销(空间)/MB | 系统吞吐量/% |
|---|---|---|---|---|
| A | 96.8 | 12 | 6 | 99.8 |
| B | 95.9 | 13 | 4 | 100 |
| C | 98.2 | 14 | 5 | 99.9 |
| D | 97.7 | 12 | 5 | 100 |
| E | 97.6 | 12 | 4 | 100 |

分析表 1 可知,本文系统的震后虚假信息入侵检测率均在 95%以上,系统的开销时间与空间分别在 12 ms 左右、5 MB 上下,吞吐量均高于 99%。数据表明,本文系统准确检测震后虚假信息入侵几率高、系统开销小、接收入侵信息的能力强,为突出本文系统的优势,实验进行不同系统的对比测试。

### 2.2　系统吞吐量与开销分析

采用吞吐量与开销评价本文系统检测震后虚假入侵性能。系统实际接收数据包量与发送数据包量比值为吞吐量,系统进行入侵信息检测花费的时间与空间为系统开销,此处以时间为例。记录仿真实

验中三种系统的吞吐量与开销情况，分别如图 6、图 7 所示。

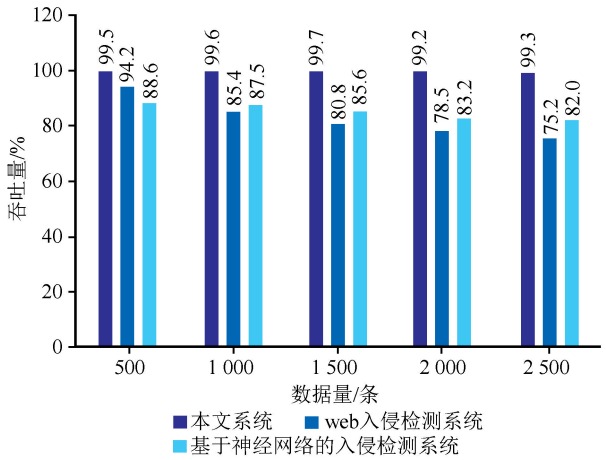


图 6 系统吞吐量对比

Fig.6 System throughput comparison

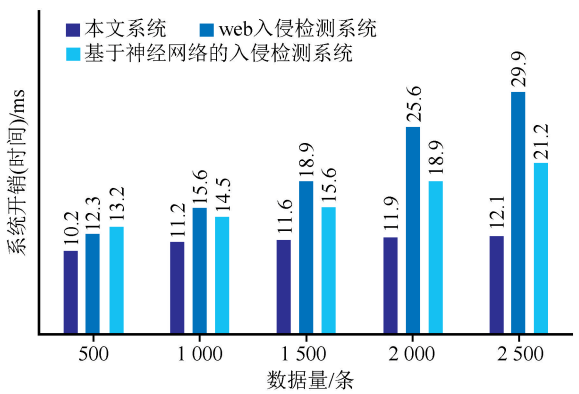


图 7 系统开销对比

Fig.7 System overhead comparison

分析图 6 能够看出，三种系统吞吐量与数据量成反比，数据量越大、系统吞吐量越小，即随着数据量的增加，系统接收的数据量有所减少。本文系统吞吐量降低趋势微弱，数据量增加到 2 500 条时，本文系统吞吐量高达 99.3%，Web 入侵检测系统与基于神经网络的入侵检测系统吞吐量分别仅为 75.2%、82%，接收到的数据量较小，降低准确检测震后虚假信息入侵的几率，相反，本文系统接收网络发送数据的能力强，降低了检测震后虚假信息入侵的难度。

图 7 数据显示，系统时间开销与数据量成正比，数据量越大、时间开销越大。本文系统接收 2 500 条数据量时花费 12.1 ms，Web 入侵检测系统花费 29.9 ms，基于神经网络的入侵检测系统花费 21.2 ms。由此可知，本文系统接收数据信息能力强、时间开销低，具有高效率的优势。

本文系统较优的数据接收能力与效率，为进行震后虚假信息实时检测提供全面的数据基础，减少入侵检测用时。

### 2.3 误报率分析

采用三种系统同时检测网络中震后虚假信息入侵情况，记录三种系统检测震后虚假入侵信息的误报率情况，如图 8 所示。

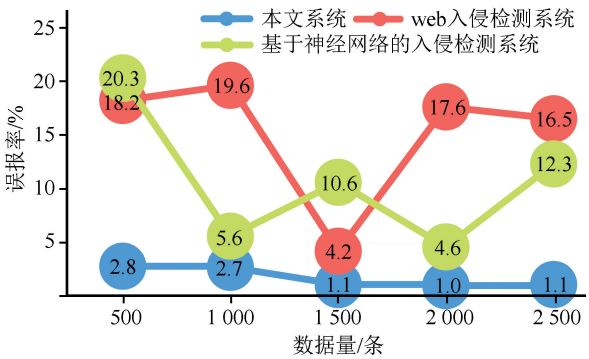


图 8 不同系统误报率对比

Fig.8 Comparison of false alarm rates in different systems

由图 8 可知，本文系统的误报率均低于 3%，检测系统进入工作状态后，逐渐适应虚假信息检测环境，误报率维持在 1%左右，基本稳定；Web 入侵检测系统大部分丢包率在 15%以上，仅在检测 1 500 条数据信息时出现丢包率最小值为 4.2%，该数据表明 Web 入侵检测系统的稳定性较差，不符合高可靠性要求；同理基于神经网络的入侵检测系统的稳定性较差、丢包率较高，分布在 4.6～20.3%之间。

## 3 结论

本文设计一种有效实时检测震后虚假信息入侵的系统。首先设计了 Web 服务器、数据包捕获模块、协议解析模块。Web 服务器包括 Web 浏览器与 Web 设备，TCP/IP 协议、HTTP 协议是连接 Web 设备与 Web 浏览器的纽带；数据包捕获模块采用 WinPcap 数据包捕获技术，具备优秀的数据包捕获机制与信息过滤机制，降低系统检测虚假信息的难度；协议解析模块主要负责捕获数据包的解码工作，解码后的数据包接口与检测系统匹配。其次，通过构建 SVM 模型将震后信息分成虚假与真实两类，根据分类结果完成震后虚假信息入侵检测。系统模块与 SVM 分类模型结合，发挥各自优势。经验证，本文系统吞吐量高达 99.3%，时间开销为 12.1 ms，优于同类型系统；本文系统检测震后虚假信息入侵的误报率在 3%以下，是一种可靠的虚假

信息入侵检测系统，其应用前景广阔。

## 参考文献(References)


[1] 张方浩，和仕芳，吕佳丽，等.基于互联网的地震灾情信息分类编码与初步应用研究[J].地震研究，2016，39(4)：664-672.
ZHANG Fanghao，HE Shifang，LYU Jiali，et al.Classification and Coding of the Earthquake-disaster Information Based on the Internet and Their Preliminary Application[J].Journal of Seismological Research，2016，39(4)：664-672.

[2] 肖健，侯建民.地震速报微博发布系统的研发[J].中国地震，2015，31(2)：456-460.
XIAO Jian，HOU Jianmin.The Development of Earthquake Information Release on WEIBO Publishing System[J].Earthquake Research in China，2015，31(2)：456-460.

[3] 张伟，巢翌，甘志强，等.结合特征分析和 Svm 优化的 Web 入侵检测系统[J].计算机仿真，2018，35(5)：406-409，447.
ZHANG Wei，CHAO Yi，GAN Zhiqiang，et al.Design of Web Intrusion Detection System Based on Feature Analysis and SVM Algorithm Optimization[J].Computer Simulation，2018，35(5)：406-409，447.

[4] 罗俊松.基于神经网络的 BP 算法研究及在网络入侵检测中的应用[J].现代电子技术，2017，40(11)：91-94.
LUO Junsong.Research on BP Algorithm Based on Neural Network and Its Application in Network Intrusion Detection[J].Modern Electronics Technique，2017，40(11)：91-94.

[5] 霍世敏，赵菊敏，李灯熬，等.针对异常序列检测的非法入侵识别算法[J].计算机工程与应用，2017，53(20)：68-74.
HUO Shimin，ZHAO Jumin，LI Deng'ao，et al.Illegal Intrusion Detection Algorithm Based on Abnormal Signal Sequences[J].Computer Engineering and Applications，2017，53(20)：68-74.

[6] AMBUSAIDI M A，HE X J，NANDA P，et al.Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm[J].IEEE Transactions on Computers，2016，65(10)：2986-2998.

[7] 赵宁，谢淑翠.基于 dpdk 的高效数据包捕获技术分析与应用[J].计算机工程与科学，2016，38(11)：2209-2215.
ZHAO Ning，XIEShucui.Analysis and Application of the High Performance Data Packet Capture Technology Based on Dpdk[J].Computer Engineering and Science，2016，38(11)：2209-2215.

[8] 张永良，张智勤，吴鸿韬，等.基于改进卷积神经网络的周界入侵检测方法[J].计算机科学，2017，44(3)：182-186.
ZHANG Yongliang，ZHANG Zhiqin，WU Hongtao，et al.Perimeter Intrusion Detection Based on Improved Convolution Neural Networks[J].Computer Science，2017，44(3)：182-186.

[9] RATHORE M M，AHMAD A，PAUL A.Real Time Intrusion Detection System for Ultra-high-speed Big Data Environments[J].The Journal of Supercomputing，2016，72(9)：3489-3510.

[10] 崔涛，刘孝刚，姜珊珊，等.变电站直流电源系统故障监测装置的研制与应用[J].江苏电机工程，2016，35(2)：26-30，38.
UI Tao，LIUXiaogang，JIANG Shanshan，et al.The Development and Application of Fault Detection for DC Power System[J].Jiangsu Electrical Engineering，2016，35(2)：26-30，38.

[11] 李军华.分层网络伪装入侵实时高速检测方法研究[J].微电子学与计算机，2017，34(10)：123-126.
LI Junhua.The Layered Network Invasion of High-Speed Real-Time Detection Method Research of Disguise[J].Microelectronics & Computer，2017，34(10)：123-126.

[12] YOON M，MOHAN，CHOI J，et al.Intrusion Detection Using Execution Contexts Learned from System Call Distributions of Real-Time Embedded Systems[J].Acta Mathematica，2015，42(1)：349-355.

[13] 马媛媛，周诚，李千目.关于网络中入侵节点信息优化检测仿真研究[J].计算机仿真，2017，34(8)：315-318.
MA Yuanyuan，ZHOU Cheng，LI Qianmu.Simulation Research on Optimization of Intrusion Node Information in Network[J].Computer Simulation，2017，34(8)：315-318.

[14] 刘云，向婵，王海花.基于互信息的特征选择在入侵检测中的优化[J].西北大学学报(自然科学版)，2017，47(5)：666-673.
LIU Yun，XIANG Chan，WANGHaihua.Optimization of Feature Selection Based on Mutual Information in Intrusion Detection[J].Journal of Northwest University (Natural Science Edition)，2017，47(5)：666-673.

[15] 杨浩，章玲玲，熊焕东，等.基于稀疏向量距离的网络入侵数据检测[J].科学技术与工程，2017，17(27)：88-92.
YANG Hao，ZHANG Lingling，XIONG Huandong，et al.Network Intrusion Detection Dased on Sparse Vector Distance[J].Science Technology and Engineering，2017，17(27)：88-92.

[16] 苏玉龙，张著洪.基于关键词的文本向量化与分类算法研究[J].贵州大学学报(自然科学版)，2018，35(3)：101-105.
SU Yulong，ZHANGZhuhong.Study on Keyword-Based Text Vectorization and Classification Algorithm[J].Journal of Guizhou University(Natural Sciences)，2018，35(3)：101-105.

[17] 崔会芳，周梦妮，王彬，等.基于格兰杰因果分析的 MCI 脑网络分类研究[J].太原理工大学学报，2018，49(6)：853-860.
CUI Huifang，ZHOU Mengni，WANG Bin，et al.Classifications of MCI Brain Network Based on Granger Causality Analysis[J].Journal of Taiyuan University of Technology，2018，49(6)：853-860.

[18] 刘燕萍，董伟，黄毕双，等.基于两阶段支持向量机的群体建筑物震害预测方法[J].华南地震，2016，36(2)：107-113.
LIU Yanping，DONG Wei，HUANG Bishuang，et al.Seismic Damage Prediction Method of Building Groups Based on a Two-stage Support Vector Machine[J].South China Journal of Seismology，2016，36(2)：107-113.