DOI: 10.13718/j.cnki.xdzk.2022.08.022

基于投票 ELM 和黑洞优化的云计算 DDoS 攻击检测

王飞雪[1]，　戴蓉[2]

1. 重庆人文科技学院 计算机工程学院，重庆 合川 401524；2. 中国民用航空飞行学院 计算机学院，四川 广汉 618307

摘要：为了解决分布式拒绝服务(Distributed Denial of Service，DDoS)攻击使云计算的最终用户无法访问云服务的问题，该文提出一种基于投票极限学习机(Voting Extreme Learning Machine，V-ELM)和黑洞优化的云计算 DDoS 攻击检测算法. 该算法采用 V-ELM 作为分类器进行系统设计，使用多个极端学习机器同时检测攻击. 使用数据包分析器捕获网络流量生成供分类器使用的样本，然后使用黑洞优化训练 V-ELM 中的所有 ELM，在攻击检测过程中将样本应用于每个 ELM 并计算输出，最后在多数投票的基础上合并得到最终输出. 实验结果表明：该文提出的算法在网络安全实验知识发现与数据挖掘(Network Security Lab Knowledge Discovery and Data Mining，NSL KDD)数据集和 KDD 分布式拒绝服务(KDD Distributed Denial of Service，KDD DDoS)数据集上的准确性、灵敏度和特异性均优于所对比的方法.

关　键　词：云计算；投票极限学习机；分布式拒绝服务攻击；黑洞优化

中图分类号：TP391　　**文献标志码**：A　　开放科学(资源服务)标识码(OSID)：

文 章 编 号：1673－9868(2022)08－0205－11

Detection of DDoS Attackin Cloud Computing Based on Voting ELM and Black Hole Optimization

WANG Feixue[1]，　DAI Rong[2]

1. *School of Computer Engineering，Chongqing College of Humanities，Science and Technology，Hechuan Chongqing 401524，China*；
2. *School of Computer Science，Civil Aviation Flight University of China，Guanghan Sichuan 618307，China*

Abstract：To solve the problem that Distributed Denial of Service (DDoS) attacks make end users of cloud computing unable to access cloud services，analgorithm fordetection of cloud computing DDoS attackbased on Voting Extreme Learning Machine (V-ELM) and black hole optimization is proposed. This algorithm

收稿日期：2021－09－18

基金项目：民航飞行技术与飞行安全重点实验室项目(FZ2020ZZ02)；国家自然科学基金民航联合基金项目(U2033213)；国家重点实验室项目(HTL-0-19K01).

作者简介：王飞雪，硕士，讲师，主要从事计算机软件与应用技术的研究.

uses V-ELM as classifier for system design, and uses multiple extreme learning machines to detect attacks at the same time. It uses packet analyzer to capture network traffic to generate samples for classifier, then uses black hole optimization to train all ELMs in V-ELM. The samples are appliedto each ELM in attack detection process and output is calculated, and finally, by combining those outputs based on the majority votes, the final output is obtained. The results show that the accuracy, sensitivity and specificity of the proposed algorithm on NSL-KDD and KDD DDoS intrusion detection dataset are better than those of the compared methods.

Key words: cloud computing; voting extreme learning machine; distributed denial of service attack; black hole ptimization

云计算是一种基于Internet技术，可按需向用户提供各种类型资源[1]，为虚拟化互联网技术(Internet Technology, IT)解决方案提供新的消费和支付的模型. 在云环境中，用户数据和应用程序分散在远程数据中心，无论用户的位置和时间如何，都可以以虚拟形式在用户端对其进行访问. 云计算以3种服务的形式提供资源：软件即服务(Software as a service, SaaS)、平台即服务(Platform as a service, PaaS)和基础架构即服务(Infrastructure as a service, IaaS). SaaS提供的资源是应用程序，PaaS提供的资源是用于开发应用程序的库和编程语言，IaaS提供的资源是处理能力、存储能力和网络带宽. 根据用户的不同，云计算基础架构可以通过4种方式进行部署：私有云、社区云、公共云和混合云[2]. 云计算由于具有按需服务、按需付费、低成本和支撑缩放等特性，正逐渐取代传统的IT实现[3]. 自动缩放意味着所提供的资源可以根据用户的需要立即放大或缩小，何时向上扩展，何时向下扩展，取决于云用户和云提供商之间的服务级别协议. 云环境中按需资源可用性的优势伴随着一些安全问题. 随着用户机密数据存储和应用程序部署在云服务提供商的端部，持续服务可用性和敏感数据保护成为云计算中面临的主要安全问题[4].

云服务使得各种客户端可以远程部署基于web技术的应用程序。但是云服务的普遍使用也存在容易受到攻击和可靠性等问题. 云计算中的服务不可用和连接性问题会完全中断服务，给消费者带来巨大的商业损失. 分布式拒绝服务(Distributed Denial of Service, DDoS)攻击[5]是针对云服务可用性的主要攻击之一. 在DDoS攻击中，攻击者在网络中搜索称为处理程序的易受攻击主机，然后在这些主机上安装恶意程序[6]. 处理程序会找到其他一些名为bots的主机，并在这些主机上安装相同的恶意程序，攻击者通过使用命令和控制机制来控制处理程序和机器人. DDoS攻击会导致云资源耗尽，云提供的服务不可用，DDoS攻击在云计算中比传统的基于基础设施的系统更危险，因为攻击者可以使用自动扩展等功能来获取利益[7]. 当云服务器上的负载增加时，云会自动向服务器分配额外的资源，称为自动扩展. 云基础架构是多租户的，在DDoS攻击期间会通过自动扩展为虚拟机提供更多资源，导致在同一物理服务器上运行的其他虚拟机的资源短缺[8].

DDoS攻击可分为基于网络层和基于应用层两类. 基于网络层的攻击使用网络层和传输层协议进行洪泛，即传输控制协议(Transmission Control Protocol, TCP)和用户数据报协议 (User Datagram Protocol, UDP)和网际互连协议(Internet Protocol, IP)；基于应用层的攻击使用变形的可扩展的标记性语言(eXtensible Markup Language,XML)消息和超文本传输协议(Hyper Text Transfer Protocol,HTTP)洪水[9]. 目前，DDoS攻击的目标是云服务器，而不是传统的服务器，DDoS攻击造成的经济损失更为危险[10]. 针对DDoS攻击的解决方案主要分为主动和被动两类. 主动式解决方案可在攻击发生之前使用各种技术区分合法用户和攻击者，合法用户在访问资源之前必须经过防御过程，会影响合法用户的易用性[11]. 在被动式解决方案中，首先在攻击发生时对其进行检测，然后针对攻击采取虚拟机迁移等缓解策略，它为合法用户提供了易用性.

DDoS攻击检测已经被许多学者研究，并提出了各种各样的解决方案. 但是，这些解决方案存在检测

精度低、误报率高等问题. 文献[12]提出一种基于流量特征和动态阈值的 DDoS 攻击检测算法，该算法提取不同的流量特征，根据 DDoS 攻击的特点计算 4 个交通特征，当计算出的特征值在一个时间间隔内大于阈值时会检测到攻击. 该阈值是动态的，确定该算法的阈值是一项具有挑战的任务，若网络流量异构时，则该算法无法应用. 文献[13]提出一种自适应人工免疫系统来缓解 DDoS 攻击，该方法基于构建适合被监控环境要求的分布式传感器网络，通过模拟不同的免疫反应建立隔离区，构建免疫记忆，根据人类生物防御机制的行为识别威胁并做出反应. 该算法需要较高的训练时间和计算能力，也需要大量数据进行训练以检测未知攻击. 文献[14]提出一种基于极限学习机(Extreme Learning Machine，ELM)的云计算 DDoS 攻击检测算法，该算法首先使用数据包分析器等工具连续捕获网络流量，预处理器连接到路由器，将云服务器连接到互联网，然后使用单 ELM 作为二类和多类的分类器进行流量分类. 该算法能够在很短的时间内训练，具有较高的攻击检测精度.

基于文献[14]的思路，本文提出一种基于投票极限学习机(Voting Extreme Learning Machine，V-ELM)和黑洞优化的云计算 DDoS 攻击检测算法. 文献[14]使用单个 ELM 进行攻击检测；本文使用多个 ELM 同时进行攻击检测，并使用黑洞优化训练所有 ELM，所有机器的结果都是在多数投票的基础上合并得到最终结果，多个 ELM 的使用提高了检测精度. 本文算法首先归一化符号特征值并建立训练数据库，然后使用数据包分析器捕获网络流量生成供分类器使用的样本. 其次，使用黑洞优化训练 V-ELM 中的所有 ELM，在攻击检测过程中将样本应用于每个 ELM 并计算输出. 然后进行多数投票，如果多数票支持攻击，则样本属于攻击类别，否则样本属于正常样本. 为了评估系统的性能，使用了 NSL KDD 和 KDD DDoS 数据集. 实现结果表明，本文提出的算法在准确性、灵敏度和特异性方面均优于所对比的方法.

本文的主要贡献：① 提出了一种基于投票极限学习机(V-ELM)和黑洞优化的云计算 DDoS 攻击检测系统；② 使用黑洞优化训练 V-ELM 中的所有 ELM；③ 使用多个极端学习机器同时检测攻击；④ 使用 NSL KDD 和 KDD DDoS 数据集.

1　准备工作

1.1　极限学习机

极限学习机(Extreme Learning Machine，ELM) 又称超限学习机，是一种具有单隐层的前馈神经网络(single hidden layer feedforward neural network，SLFN). 极限学习机框图如图 1 所示.

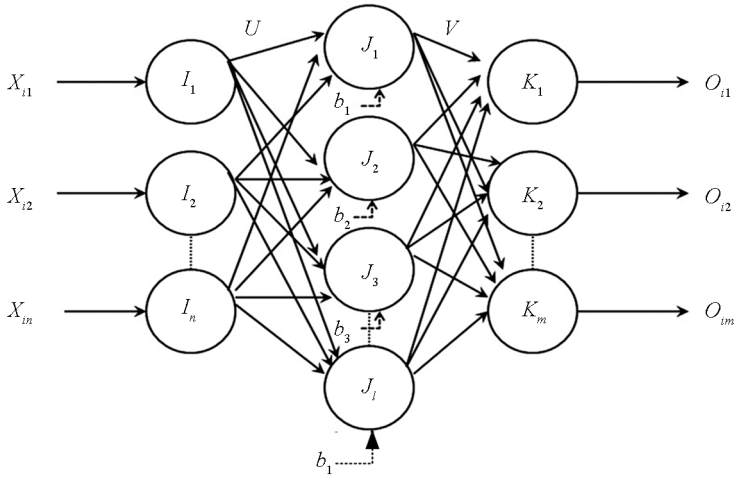

图 1　极限学习机框图

在 ELM 中，输入层与隐藏层之间的权值和隐藏层偏差是随机初始化的. 设一个 ELM 有 l 个隐藏的神经元，在隐藏层中有激活函数 f，如果存在 N 个形式为($\boldsymbol{x}_i$，$\boldsymbol{t}_i$) 的训练样本，则该网络的输出可以建模为：

$$\boldsymbol{o}_i = \sum_{j=1}^{l} \boldsymbol{v}_j f(\boldsymbol{u}_j, b_j, \boldsymbol{x}_i) \tag{1}$$

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})^T \in \boldsymbol{R}^n \tag{2}$$

$$\boldsymbol{t}_i = (t_{i1}, t_{i2}, \cdots, t_{im})^T \in \boldsymbol{R}^m \tag{3}$$

$$\boldsymbol{v}_j = [v_{j1}, v_{j2}, \cdots, v_{jm}]^T \tag{4}$$

其中，$i=1,2,3,\cdots,N$，Q_i 是网络输出，$\boldsymbol{u}_j \in \boldsymbol{R}^n$ 和 $b_j \in \boldsymbol{R}$ 是第 j 个隐藏神经元的学习参数，$\boldsymbol{R}$ 表示一元有序实数组，$\boldsymbol{R}^n$ 表示 n 元有序实数组，$\boldsymbol{R}^m$ 表示 m 元有序实数组，$\boldsymbol{v}_j$ 表示连接第 j 个隐藏神经元和输出神经元的权值向量.

本文将所有 N 个样本的方程写成：

$$\boldsymbol{O} = \boldsymbol{H}\boldsymbol{V} \tag{5}$$

$$\boldsymbol{H} = \begin{bmatrix} f(\boldsymbol{u}_1, b_1, \boldsymbol{x}_1) & \cdots & f(\boldsymbol{u}_l, b_l, \boldsymbol{x}_1) \\ \vdots & \vdots & \vdots \\ f(\boldsymbol{u}_1, b_1, \boldsymbol{x}_N) & \cdots & f(\boldsymbol{u}_1, b_l, \boldsymbol{x}_N) \end{bmatrix} \tag{6}$$

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_t^T \end{bmatrix} \qquad \boldsymbol{O} = \begin{bmatrix} \boldsymbol{o}_1^T \\ \vdots \\ \boldsymbol{o}_N^T \end{bmatrix} \tag{7}$$

$$\boldsymbol{T} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_N) \tag{8}$$

其中，O 为网络输出，T 为样本的标签向量. 为了最小化误差 $\| \boldsymbol{O} - \boldsymbol{T} \|$，可以随机初始化输入隐藏权重矩阵 $\boldsymbol{u}_j$ 和隐藏偏置 b_j，通过黑洞优化来计算隐藏的输出权重矩阵.

1.2 云计算

云计算是一种基于因特网的技术，以3种服务的形式提供其资源：

1) 软件即服务(SaaS)：云服务提供商负责运行和维护应用软件、操作系统等资源. SaaS模型在客户看来是一个基于web的应用程序接口，其中因特网用于提供使用web浏览器访问的服务. 托管应用程序可以通过智能手机和笔记本电脑等不同的设备进行访问. 与传统软件不同，SaaS的优势在于客户无需在自己的计算机上购买、安装、升级、维护或运行软件.

2) 平台即服务(PaaS)：云服务提供的资源用于开发应用程序的库和编程语言. 云服务提供商提供、运行、维护系统软件(即操作系统)和其他计算资源. PaaS服务包括应用程序的设计、开发和托管. 其他服务包括协作、数据库集成、安全性、web服务集成、伸缩性等.

3) 基础架构即服务(IaaS)：云服务提供的资源包括处理能力、存储和网络带宽. 该项服务可以减少资本支出成本，用户仅为所需的服务付费，可以随时根据自己的需求扩展和缩小资源.

根据用户不同，可以使用不同类型的云计算基础架构. 云计算基础架构可以通过4种方式进行部署. ① 私有云：仅为组织用户设置；② 社区云：为两个或多个组织的用户设置，将服务和基础结构提供给具有相似兴趣的组织；③ 公共云：它是为全球用户建立的，为公众或任何组织提供基础设施和服务，资源被成百上千的人共享；④ 混合云：它由两种或多种类型的云计算基础架构组成. 这种类型的云是私有云和公共云的混合体. 虽然云是混合的，但是每个云都有自己的身份，因此有助于多个部署.

1.3 DDos攻击

在DDoS攻击中，攻击者在网络中搜索易受攻击的主机(称为处理程序)，然后在这些主机上安装恶意程序. 处理程序会找到其他一些称为bots的主机，并在这些主机上安装相同的恶意程序. 攻击者通过使用命令和控制机制来控制处理程序和机器人，最终攻击由机器人执行.

DDoS攻击大致分为两类：

1) 网络层攻击：这些攻击根据传输控制协议(Transmission Control Protocol，TCP)，用户数据报协议

(User Datagram Protocol，UDP)或互联网控制报文协议(Internet Control Message Protocol，ICMP)包进行洪泛. ICMP flood，UDP flood 和 TCP 同步(Synchronization，SYN) flood 就是这类攻击的例子.

2) 应用层攻击：包括 HTTP 洪水攻击和 XML 洪水攻击. 在 HTTP 洪水攻击中，云中的 web 服务器被扭曲的 HTTP 包淹没. XML 洪水攻击用于使用简单对象访问协议(Simple Object Access Protocol，SOAP)消息的 web 服务中.

DDoS 攻击检测和防御方法主要分为 3 类：基于特征的 DDoS 攻击检测和防御、基于异常的 DDoS 攻击检测和防御、基于特征和基于异常的混合方法的 DDoS 攻击检测和防御. DDoS 攻击检测技术分类如图 2 所示. 基于特征的技术的局限性：无法检测到新的攻击、维护更新的签名数据库非常困难、由于对信号的误解导致高假阴性率. 基于异常的技术科技进一步分为统计、机器学习[如人工神经网络(Artificial Neural Network，ANN)、支持向量机(Support Vector Machine，SVM)、K-近邻(K-Nearest Neighbor，KNN)等算法]和软件定义网络(Software Defined Network，SDN). 统计技术的局限性为确定阈值是一项具有挑战性的任务，当网络流量异构时，则无法应用. 机器学习技术的局限性需要较高的训练时间和计算能力，也需要大量数据进行训练以检测未知攻击.

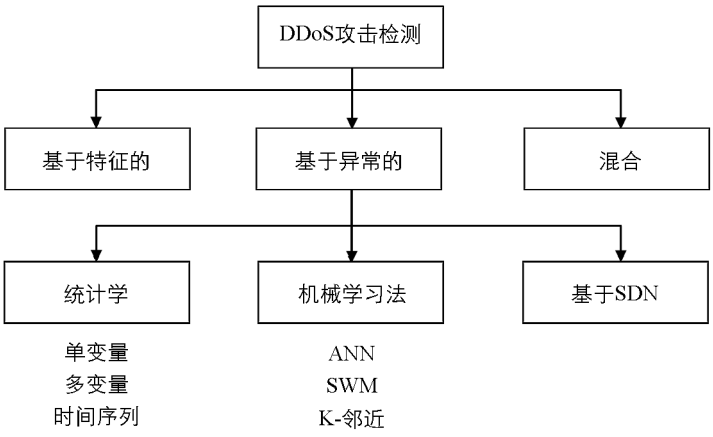

图 2　DDoS 攻击检测技术的分类

2　基于 V-ELM 和黑洞优化的云计算 DDoS 攻击检测

基于 V-ELM 和黑洞优化的云计算 DDoS 攻击检测采用投票极限学习机(ELM)作为分类器进行系统设计，使用多个 ELM 同时检测攻击. 首先使用相同的训练数据集基于黑洞优化对 V-ELM 中的所有 ELM 进行培训，在攻击检测过程中将样本应用于每个 ELM 并计算输出，然后进行多数投票. 如果多数票支持攻击，则样本属于攻击类别，否则样本属于正常样本. 本文所提算法的 DOS 攻击检测模型如图 3 所示，DOS 攻击检测具体流程如图 4 所示.

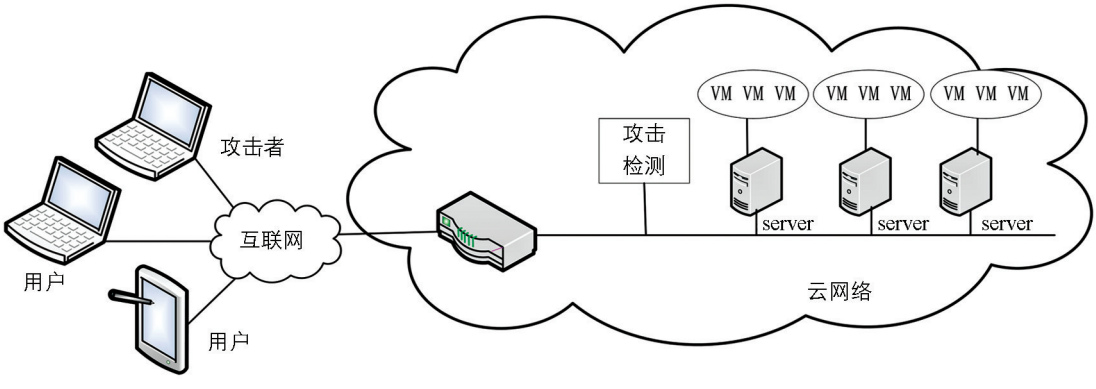

图 3　本文所提算法的 DDoS 攻击检测模型

2.1　归一化特征值

基于 V-ELM 的 DDoS 攻击检测是有监督分类器的，这意味着在使用分类器检测攻击之前，必须使用已知类别标签对分类器进行训练. 为了训练分类器，使用了一个标记样本数据库. 为了建立训练数据库，采用过去的网络流量数据，包含特定时间段内两种类型数据包(正常和恶意数据包)的信息.

首先，为每个包提取特征以及包的类型，将符号特征值替换为数值，然后执行归一化. 归一化是一个在[0，1]范围内缩放特征值的过程. 归一化对于学习非常重要，因为归一化后所有的特性都在同一个级别上，便于不同单位或量级的指标能够进行比较和加权. 使用式(1)进行归一化：

$$x_{ij}=\frac{x'_{ij}-x'_j(\min)}{x'_j(\max)-x'_j(\min)} \tag{9}$$

其中，x_{ij} 表示第 i 个样本中第 j 个特征的归一化值，x'_{ij} 表示第 i 个样本中第 j 个特征的非归一化值，$x'_j(\min)$ 表示所有未归一化样本中第 j 个特征的最小值，$x'_j(\max)$ 表示所有未归一化样本中第 j 个特征的最大值.

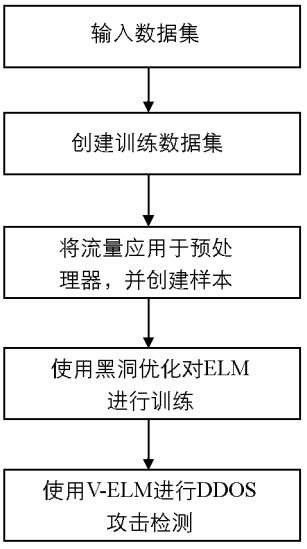

图4 本文所提算法的DDoS攻击检测流程图

2.2 创建训练数据集

训练数据集是具有($\boldsymbol{x}_i$，$\boldsymbol{t}_i$)形式的 N 个样本的数据集. $\boldsymbol{x}_i=[x_{i1}, x_{i2}, x_{i3}, \cdots, x_{im}]$ 是一个长度为 m 的第 i 个样本的所有特征向量，这些特征用于将样本分类为攻击样本或正常样本. $\boldsymbol{t}_i=[1, 0]$ 或 $\boldsymbol{t}_i=[0, 1]$ 是第 i 个样本(攻击样本或正常样本)的标签向量或类型向量.

在本文中，$\boldsymbol{t}_i=[1, 0]$ 表示第 i 个样本属于攻击样本类，$\boldsymbol{t}_i=[0, 1]$ 表示第 i 个样本属于正常样本类. 假设示例由4个特征字符化：源IP、目标IP、源端口和目标端口，则 $\boldsymbol{x}_i$ 的格式为 $\boldsymbol{x}_i=$[源IP，目标IP，源端口，目标端口]，$\boldsymbol{t}_i$ 的格式为 $\boldsymbol{t}_i=[1, 0]$ 或 $\boldsymbol{t}_i=[0, 1]$. 样本($\boldsymbol{x}_1$，$\boldsymbol{t}_1$)=([200.10.1.2，10.23.11，5 000，80]，[0，1]) 表示它属于正常样本类.

2.3 预处理器捕获网络流量和生成样本

预处理器连接路由器，路由器将云服务器连接到因特网. 预处理器的工作是使用流量捕获工具不断地对网络流量进行捕获，并生成样本供分类器使用.

本文使用Wireshark数据包分析器等工具连续捕获网络流量，样品分组生成，使用每个时段 t 期间的捕获流量. 在每个时段 t 期间捕获的数据保存在单独的文件中，对于每个文件创建一组示例. 从每个文件中选择与培训数据库样本相同的特征，将符号特征值替换为数值，并使用式(1)进行归一化，使特征值在[0，1]范围内缩放. 然后，构造一组形式为 $\boldsymbol{y}_i=[y_{i1}, y_{i2}, \cdots, y_{in}]$ 的样本，其中 $\boldsymbol{y}_i$ 的所有特征都与 $\boldsymbol{x}_i$ 的特征相同，并将这组样本应用于分类器.

2.4 基于黑洞优化对每个ELM进行培训

为了提高检测的准确性，使用多个ELM进行分类，每个ELM都使用黑洞优化进行培训. 黑洞优化是一种基于种群的算法，可以最小化或最大化目标函数. 该算法首先生成解的总体，然后用每个解计算目标函数. 一个解表示ELM中使用的所有权重的向量. 目标函数最小化由式(10)给出均方误差，然后选择提供最小均方误差输出的权重向量.

$$f_{obj}=\min\left(\frac{1}{N}\sum_{i=1}^{N}(T_i-O_i)^2\right) \tag{10}$$

其中，N 是训练数据库中的样本总数，T_i 和 O_i 是第 i 个样本的实际目标和ELM输出.

目标函数最小值的解称为黑洞，其余的解称为恒星. 黑洞吸引其他恒星，因此它们的位置发生了变化. 新的位置由式(11)计算：

$$pos_i(t+1)=pos_i(t)+r\times(pos_{BH}-pos_i(t)) \tag{11}$$

其中，$pos_i(t+1)$ 和 $pos_i(t)$ 分别代表第 i 星的新旧位置，pos_{BH} 代表黑洞的位置，r 是[0，1]中的随机数. $pos_i(t+1)$，$pos_i(t)$ 和 pos_{BH} 均为权重向量.

如果一颗新位置的恒星具有比黑洞更好的目标函数，那么它的位置就会与黑洞交换. 如果一颗恒星越过黑洞的视界半径，它就会被黑洞吞噬，并产生另一颗随机位置的恒星. 黑洞视界半径和恒星距黑洞的距离由式(12) 计算.

$$Rad=\frac{f_{BH}}{\sum_{i=1}^{N}f_i},\ D_i=f_{BH}-f_i \tag{12}$$

其中，R_{ad} 表示黑洞视界半径，D_i 表示恒星距黑洞的距离，f_i 表示恒星的目标函数值，f_{BH} 表示黑洞的目标函数值. 该分类器经过训练，可用于检测攻击. 它接受来自预处理器 $\boldsymbol{y}_i=[y_{i1}, y_{i2}, \cdots, y_{in}]$ 形式的输入并生成输出 $\boldsymbol{O}_i=[1, 0]$ 或 $\boldsymbol{O}_i=[0, 1]$. $\boldsymbol{O}_i=[1, 0]$ 表示这是一个攻击样本，$\boldsymbol{O}_i=[0, 1]$ 表示正常样本.

2.5　基于 V-ELM 的 DDoS 攻击检测

在 V-ELM 中，使用 k 个单独的 ELM，这些单独的 ELM 用相同数量的隐藏神经元(l) 初始化. 随机初始化每个 ELM 的学习参数 $\boldsymbol{u}_j$ 和 b_j($j=1, 2, \cdots, l$). 然后，使用黑洞优化进行训练，在分类过程中将样品应用到每个 ELM 上，并计算输出，最后根据多数投票计算得出最终输出(图 5). 在分类阶段，一组形式为 $\boldsymbol{y}_i=[y_{i1}, y_{i2}, \cdots, y_{in}]$ 的样本适用于所有 ELM. 取每个 ELM 的输出 $\boldsymbol{O}_i=[1, 0]$ 或 $\boldsymbol{O}_i=[0, 1]$，并使用多数投票获得最终输出.

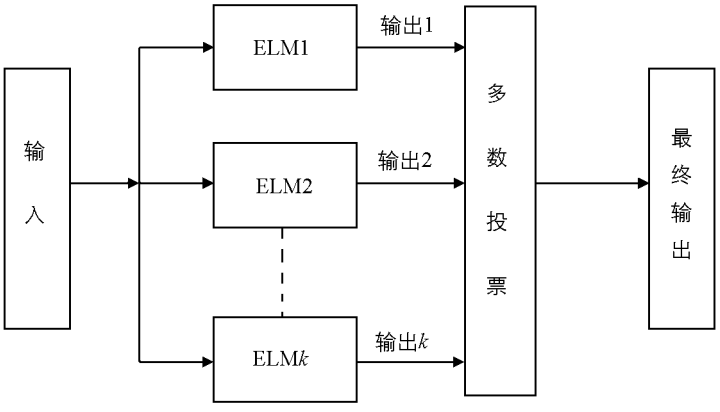

图 5　V-ELM 的工作原理

对于每个样本 i，使用向量 $\boldsymbol{C}_i \in \boldsymbol{Z}^2_{>=0}$ 来存储由不同 ELM 标识的类标签. 最初，$\boldsymbol{C}_i$ 初始化为零($\boldsymbol{C}_i=[0, 0]$). 当发生 ELM 输出时，$\boldsymbol{C}_i$ 中相应位置的值增加 1. 例如，假设 ELM_1 的输出为[1, 0]，那么 $\boldsymbol{C}_i$ 变为[1, 0]. ELM_2 的输出再次为[1, 0]，现在 $\boldsymbol{C}_i$ 变为[2, 0]. ELM_3 的输出为[0, 1]，现在 $\boldsymbol{C}_i$ 变为[2, 1].

当所有 k 个输出到达时，最终输出计算如下：首先计算最大值 $\max(\boldsymbol{C}_i)$，然后计算 $\max(\boldsymbol{C}_i)$ 的位置. 然后根据 $pos[\max(C_i)]$ 进行样本判别，$pos[\max(C_i)]$ 有两种状态，本文设置为 1 和 2. 如果，$pos[\max(\boldsymbol{C}_i)]=1$，则为攻击样本；$pos[\max(\boldsymbol{C}_i)]=2$，则为正常样本. 如果发现攻击示例，则向云管理员生成警报.

3　实验结果与分析

为了评估本文所提算法的性能，所有实验均在一台配置为 Intel Core i7 CPU @3.60Ghz 和 8 GB RAM 的机器上进行，所有测试均在 Matlab 2014a 环境下实现. 选取 NSL KDD 和 KDD DDoS 数据集对算法性能进行评估，NSL KDD 数据集包含几种类型的异常数据包以及正常数据包；KDD DDoS 数据集包含 DDoS 数据包以及正常数据包. NSL KDD 数据集有 5 755 个网络流量实例，其中包括 2 152 个正常网络流量实例和 3 603 个攻击流量实例. 由于数据集的异常实例不属于 DDoS 类别，在 NSL KDD 数据集中选择出最相关的特征，并使用选定的特征从数据集中分离出 DDoS 数据包，将所选的 DDoS 数据包和正常数据包组成 KDD DDoS 数据集. KDD DDoS 数据集使用怀卡托智能分析环境(Waikato Environment for Knowledge Analysis，WEKA)中的离散滤波器工具进行离散化处理获得更好的性能. 表 1 给出了有关数据集更多的详细信息，表 2 给出了算法的仿真参数. 使用 S 型函数作为激活函数，其定义为：

$$Sigmoid(x)=\frac{1}{1+e^{-x}} \tag{13}$$

其中，x 表示常数，e 表示指数.

表1 数据集信息

数据集名称	NSL KDD	数据集名称	NSL KDD
特征数量	41	测试样本	5 000
训练样本	20 000		

表2 仿真参数

数据集名称	NSL KDD	KDD DDoS
隐藏神经元数量(l)	1 000	60
激活函数(f)	S型函数	S型函数
迭代数	1 000	1 000
初始人口(对于黑洞 ANN)	32	32
ELM 数量(k)	36	36

3.1 评价指标

为了评估算法性能，采用3种指标进行评估：准确度(Accuracy，Acc)、灵敏度(Sensitivity，Sen)和特异性(specificity，$Spec$).

准确度(Acc)表示攻击检测系统准确分类的数据比例，定义为：

$$Acc=\frac{TP+TN}{TP+FP+TN+FN}\times 100\% \tag{14}$$

灵敏度(Sen)表示在所有攻击中检测到的攻击所占的比例，定义为：

$$Sen=\frac{TP}{TP+FN}\times 100\% \tag{15}$$

特异性($Spec$)表示在所有正常中检测到的正常所占的比例，定义为：

$$Spec=\frac{TN}{FP+TN}\times 100\% \tag{16}$$

其中，真阳性(True Positive，TP)定义为属于攻击并由分类器正确识别的样本数；真阴性(True Negative，TN)定义为属于正常且由分类器正确识别的样本数；假阳性(False Positive，FP)定义为属于正常但被分类器错误识别为攻击的样本数；假阴性(False Negative，FN)定义为属于攻击但被分类器错误地识别为正常的样本数.

3.2 实验结果与分析

图6给出了本文所提算法在NSL KDD数据集和KDD DDoS数据集均进行50次实验的平均检测精度，具体实验数据如表3所示.

表3 50次实验的检测精度 %

数据集名称	实验次数				
	10	20	30	40	50
NSL KDD	99.14	99.24	99.14	99.24	99.24
KDD DDoS	90.66	92.58	92.08	92.23	92.54

从图6中可以看出，本文提出的算法在NSL KDD数据集下具有较高的检测精度，并且NSL KDD数

据集在精度上的差异几乎可以忽略不计. 这是因为使用黑洞优化训练投票极限学习机中的所有 ELM，并使用多个 ELM 同时进行攻击检测，然后将所有机器的结果在多数投票的基础上合并得到最终结果.

图 7 给出了在保持隐藏神经元(l)数量不变(在 NSL KDD 数据集中 $l=1\,000$，在 KDD DDoS 数据集中 $l=60$)的情况下，ELM 的数量变化对检测精度的影响情况. 具体实验数据如表 4 所示.

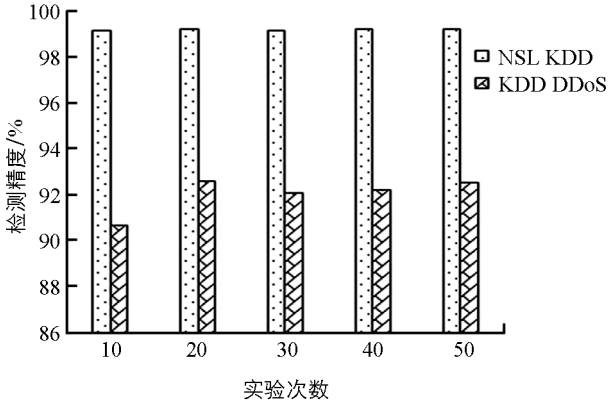

图 6　实验的检测精度

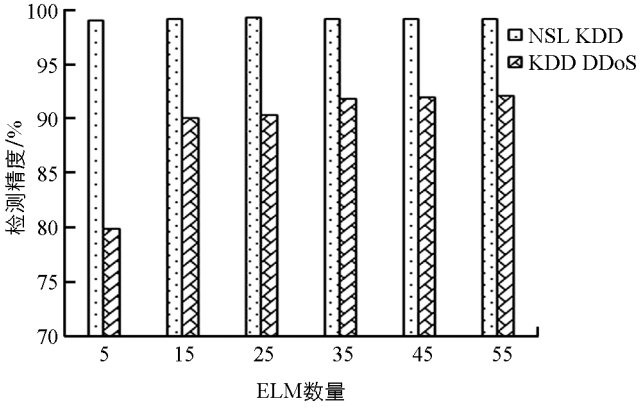

图 7　精度随 ELM 的变化情况

表 4　精度随 ELM 数量的变化情况　%

数据集名称	ELM 数量					
	5	15	25	35	45	55
NSL KDD	98.96	99.20	99.30	99.18	99.07	99.18
KDD DDoS	79.86	90.05	90.40	91.85	92.01	92.13

从图 6、图 7 中可以看出，本文提出的算法在 KDD DDoS 数据集上具有较高的检测精度，使用 NSL KDD 数据集时精度几乎没有变化；在 KDD DDoS 数据集中，当 ELM 数量小于 35 时，精度随着 ELM 数量的增加而增加，ELM 数量大于 35 时，精度没有显著提高.

表 5、表 6、图 8 和图 9 给出了在 NSL KDD 数据集和 KDD DDoS 数据集中，本文算法投票极限学习机(V-ELM)、极限学习机(ELM)[14]和随机森林(RF)[15]的性能指标. 由图 8、图 9 可以看出，与其他算法相比，本文算法在 NSL KDD 数据集和 KDD DDoS 数据集的准确性、灵敏度和特异性均优于所对比的方法.

表 5　各算法在 NSL-KDD 数据集的性能指标　%

性能	算法		
	ELM	RF	V-ELM
准确性	98.90	98.95	99.18
灵敏度	99.40	99.46	99.50
特异性	98.46	98.84	98.86

表 6　各算法在 KDD DDoS 数据集的性能指标　%

性能	算法		
	ELM	RF	V-ELM
准确性	90.72	91.54	92.11
灵敏度	81.54	85.11	85.34
特异性	98.75	99.31	99.77

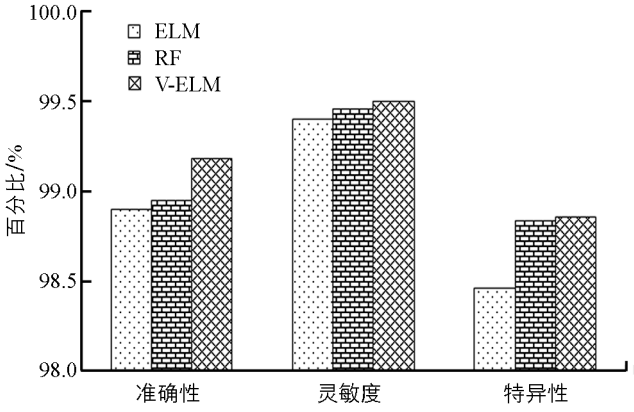

图8 各算法在NSL KDD数据集的性能指标

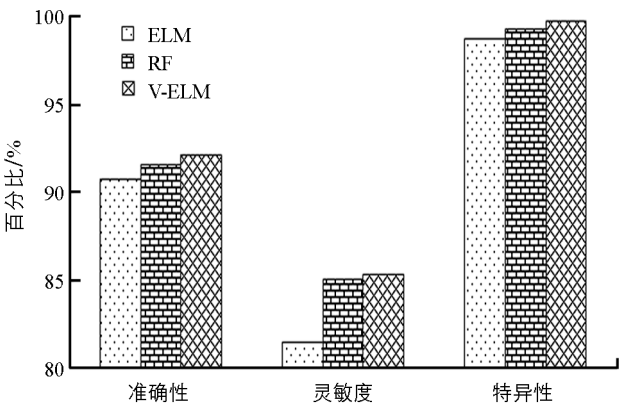

图9 各算法在KDD DDoS数据集的性能指标

表7给出了在NSL KDD数据集和KDD DDoS数据集中各算法的训练时间. 由表7可以看出，本文算法在NSL KDD数据集和KDD DDoS数据集上的执行时间比其他方法长，这是因为ELM算法使用的是单个网络进行攻击检测；RF算法对训练样本随机有放回地抽取，产生多个训练数据的子集，然后构造分类器进行攻击检测；而本文使用多个ELM同时进行攻击检测，最后根据多数投票计算得出最终输出.

表7 各算法的训练时间 s

数据集名称	训练时间		
	ELM	RF	V-ELM
NSL KDD	4.34	7.7	141.57
KDD DDoS	0.46	3.16	2.76

图10显示在NSL KDD数据集下，同时改变ELM数量(k)和隐层神经元数量(l)的攻击检测准确率. 准确率通过颜色变化来描述. 左上角表示最低的准确率(96.84%，)，右下角表示最高的准确率(99.20%). 在NSL KDD数据集情况下，准确率随 k 和 l 值的增加而增加.

图11显示在KDD DDoS数据集下，同时改变ELM数量(k)和隐层神经元数量(l)的攻击检测准确率. 准确率通过颜色变化来描述. 对于KDD DDoS数据集，初始准确率随 k 和 l 值的增加而增加，当增加到最大值后，准确率会随着 k 和 l 值的增加而减小.

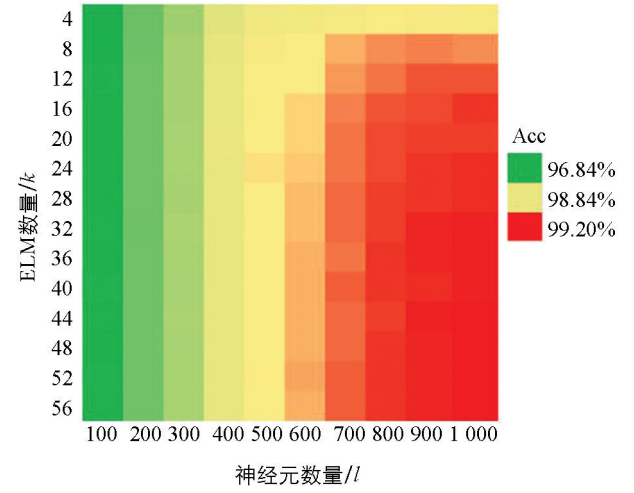

图10 不同 k 和 l 值下NSL KDD的平均检测准确率

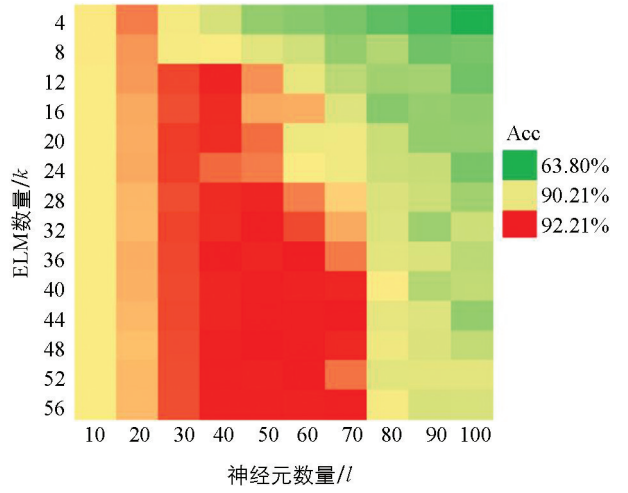

图11 不同 k 和 l 值下KDD DDoS的平均检测精度

4 结论

为了安全地使用云计算，本文提出一种基于V-ELM和黑洞优化的DDoS攻击检测算法. 该算法归一

化符号特征值并建立训练数据库，使用数据包分析器捕获网络流量生成供分类器使用的样本. 使用黑洞优化训练投票极限学习机中的所有 ELM，并使用多个 ELM 同时进行攻击检测，得到最终结果. 如果多数票支持攻击，则样本属于攻击类别，否则样本属于正常样本. 为了评估系统的性能，本文使用了 NSL-KDD 数据集和 KDD DDoS 数据集进行测试. 实验结果表明，本文提出的算法在准确性、灵敏度和特异性方面均优于所对比的方法. 未来的工作是考虑在多个入侵检测数据集上进行实验，从而更全面地验证本文所提算法的效果.

参考文献：

[1] DIZDAREVIĆ J, CARPIO F, JUKAN A, et al. A Survey of Communication Protocols for Internet of Things and Related Challenges of Fog and Cloud Computing Integration [J]. ACM Computing Surveys, 2019, 51(6): 1-29.

[2] VARGHESE B, BUYYA R. Next Generation Cloud Computing: New Trends and Research Directions [J]. Future Generation Computer Systems, 2018, 79: 849-861.

[3] STERGIOU C, PSANNIS K E, KIM B G, et al. Secure Integration of IoT and Cloud Computing [J]. Future Generation Computer Systems, 2018, 78: 964-975.

[4] DARWISH A, HASSANIEN A E, ELHOSENY M, et al. The Impact of the Hybrid Platform of Internet of Things and Cloud Computing on Healthcare Systems: Opportunities, Challenges, and Open Problems [J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(10): 4151-4166.

[5] 王磊，李刚，王斐玉. 改进属性加密结合代理重加密的云计算安全访问控制策略 [J]. 计算机应用与软件，2019，36(7)：327-333.

[6] TCHERNYKH A, SCHWIEGELSOHN U, TALBI E G, et al. Towards Understanding Uncertainty in Cloud Computing with Risks of Confidentiality, Integrity, and Availability [J]. Journal of Computational Science, 2019, 36: 100581.

[7] BHUSHAN K, GUPTA B B. Distributed Denial of Service (DDoS) Attack Mitigation in Software Defined Network (SDN)-Based Cloud Computing Environment [J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(5): 1985-1997.

[8] 于鹏程，戚湧，李千目. 基于随机森林分类模型的 DDoS 攻击检测方法 [J]. 计算机应用研究，2017，34(10)：3068-3072.

[9] SAXENA R, DEY S. DDoS Attack Prevention Using Collaborative Approach for Cloud Computing [J]. Cluster Computing, 2020, 23(2): 1329-1344.

[10] Alarqan M A, Zaaba Z F, Almomani A. Detection Mechanisms of DDoS Attack in Cloud Computing Environment: A Survey [C] //International Conference on Advances in Cyber Security. Penang: Springer, 2019.

[11] AGRAWAL N, TAPASWI S. Defense Mechanisms Against DDoS Attacks in a Cloud Computing Environment: State-of-the-Art and Research Challenges [J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3769-3795.

[12] DAVID J, THOMAS C. Efficient DDoS Flood Attack Detection Using Dynamic Thresholding on Flow-Based Network Traffic [J]. Computers & Security, 2019, 82: 284-295.

[13] VIDAL J M, OROZCO A L S, VILLALBA L J G. Adaptive Artificial Immune Networks for Mitigating DoS Flooding Attacks [J]. Swarm and Evolutionary Computation, 2018, 38: 94-108.

[14] KUSHWAH G S, ALI S T. Distributed Denial of Service Attacks Detection in Cloud Computing Using Extreme Learning Machine [J]. International Journal of Communication Networks and Distributed Systems, 2019, 23(3): 328-351.

[15] CHENG J R, LI M Y, TANG X Y, et al. Flow Correlation Degree Optimization Driven Random Forest for Detecting DDoS Attacks in Cloud Computing [J]. Security and Communication Networks, 2018, 2018: 6459326.

责任编辑　夏娟