基于双层异质集成学习器的入侵检测方法

凌 玥[1,2], 刘玉岭[1,2], 姜 波[1], 李 宁[1], 卢志刚[1,2], 刘宝旭[1,2]

[1] 中国科学院信息工程研究所, 北京 中国 100093
[2] 中国科学院大学网络空间安全学院, 北京 中国 100049

摘要 入侵检测是网络安全领域中具有挑战性和重要性的任务。现有研究以增加时间消耗和误报率为代价, 重点关注如何提高检测率, 在实际应用中代价较大。为此, 本文提出了一种使用双层异质学习器集成学习策略的入侵检测 IDHEL 模型。该模型使用概率核主成分分析方法降低数据维度, 采用多个异质分类器通过分层十折交叉验证策略进行异常检测, 并根据所提出的分类器评估算法筛选出在相关数据上表现最佳的三种分类器, 基于概率加权投票的多分类器集成算法进行入侵检测。实验结果表明 IDHEL 模型在准确率、错误率和时间消耗方面均优于现有主流入侵检测模型。

关键词 入侵检测; 异质学习器集成; 概率核主成分分析; 分类器评估; 概率加权投票
中图法分类号 TP391.1 DOI 号 10.19363/J.cnki.cn10-1380/tn.2021.05.02

Intrusion Detection Method based on Double-Layer Heterogeneous Ensemble Learner

LING Yue[1,2], LIU Yuling[1,2], JIANG Bo[1], LI Ning[1], LU Zhigang[1,2], LIU Baoxu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract Intrusion detection is a challenging and important task. Nowadays, researchers proposed many intrusion detection models and technologies. However, existing research focused on how to increase detection rate at the cost of increasing time consumption and false positive rate, which is costly in practical application. In this paper, we propose a novel intrusion detection model using double-layer heterogeneous ensemble learner strategy (IDHEL). This model first uses probabilistic kernel principal component analysis to efficiently reduce the data dimension, in order to reduce the computational overhead. Then, multiple heterogeneous classifiers are adopted for anomaly detection by a layered ten-fold cross validation strategy. Finally, IDHEL chooses the best three classifiers based on probability-weighted voting for intrusion detection. We compare the IDHEL model with existing algorithms, and the experimental results have shown that the IDHEL model is superior to other models in terms of accuracy, False Positive Rate (FPR) and time consumption.

Key words intrusion detection; heterogeneous classifiers ensemble; probabilistic kernel principal component analysis; classifier evaluation; probability weighted voting

1 引言

为了减少和防止网络攻击, 我们有必要尽可能发现各种攻击企图、攻击行为或者攻击结果, 以保证网络系统资源的完整性、机密性和可用性。传统的安全措施, 如身份认证和防火墙, 只能被动地保护网络安全, 而且对网络管理员的要求很高。实际上, 入侵检测技术是检测攻击的最常用方法, 它可以很好地解决上述问题[1]。业界最为常用的基于误用的入侵检测采用的是签名模式匹配的方法, 它通过建立正常的行为模型来监控入侵标志的网络流量信息。以下有几方面原因可以解释为什么应该采用基于异常的检测方法: 一方面, 基于误用的入侵检测技术很难检测到未写入规则库的攻击, 因此通常具有较高的误报率和漏报率; 另一方面, 在发现新攻击和部署其相应签名之间可能存在较大的时间间隔, 安

通讯作者: 卢志刚, 博士, 高级工程师, Email:luzhigang@iie.ac.cn。
本课题得到中国自然科学基金(No. 61702508，No.61802404), 国家重点研发计划课题(No. 2016YFF0204002, No. 2016YFF0204003), “十三五”装备预研领域基金(No. 6140002020115)的支持，也得到中国科学院网络评估技术重点实验室和北京市网络安全与保护技术重点实验室的部分支持。
收稿日期: 2019-05-31; 修改日期: 2019-09-16; 定稿日期: 2021-03-05

全管理员也需对开发的签名进行管理、分发、保持最新。一旦攻击者稍微修改一些已知的恶意软件, 就会带来较大挑战。因此, 在本文中, 我们关注如何提高基于异常检测技术的实用性。

在基于异常的网络入侵检测中, 很多研究者采用了机器学习算法, 并对其进行了改进和应用[2-7]。其基本流程如图 1 所示。首先是获取数据, 然后对数据进行预处理, 最主要的是对数据进行探索性分析, 主要包含了四个步骤: 特征选择/降维处理、模型选择、模型部署、模型验证与优化。

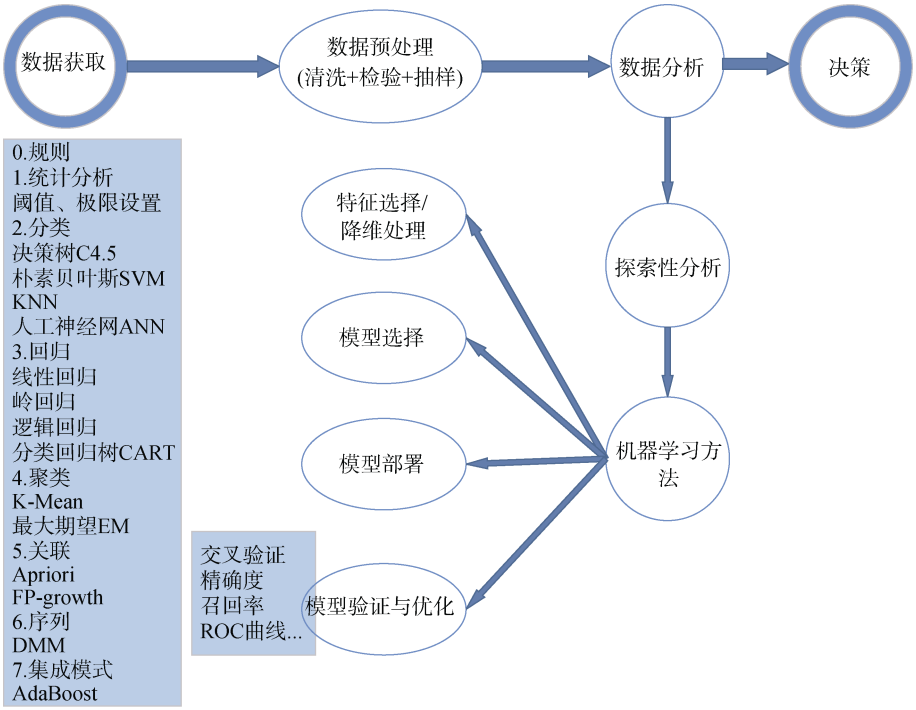

图 1　采用机器学习的入侵异常检测流程图

Figure 1　Flow chart of anomaly-based intrusion detection using machine learning method

我们将传统机器学习的入侵检测方法分为两类: 使用单个分类器[2,3,8-10]进行入侵检测, 以及融合多个分类器来进行检测[6-7]。使用单个分类器的检测方法具有较高错误率, 因为在分类过程中, 这些方法的性能通常会随着不同的分类器和/或不同的数据集的变化而变化, 因此会产生较高错误率。而以合理的方式融合多个分类器可以减少整体分类错误并增强模型的泛化能力。这个融合过程被称为集成学习。近年来, 基于深度学习的异常检测研究也越来越广泛[10-11]。然而, 由于缺乏理论基础、超参数和网络设计, 深度神经网络被认为是一个“黑匣子”, 其计算非常耗时, 解释性也较差。同时, 通过应用传统的机器学习方法, 可以轻松调整超参数并改变模型设计。因此, 使用传统的机器学习模型更具说明性和效率。集成学习具有很强的泛化能力, 可以降低错误率, 因此几种传统分类器的组合可以降低错误率, 使我们能够更全面地了解数据和底层算法。几种入侵检测方法的分析如下表所示。

表 1　入侵检测方法对比分析

Table 1　Comparative analysis of intrusion detection methods

已有方法	缺点
基于误用的入侵检测方法 (采用签名模式匹配)[12]	只适用于已知的攻击
基于异常的入侵检测方法 (改进传统机器学习算法)[3,5]	误报率、漏报率较高
基于异常的入侵检测方法 (深度学习)[10-11]	时间消耗巨大+可解释性差
基于异常的入侵检测方法 (分类器集成)[6-7]	花费大量时间调整权重

对于入侵检测, 我们还需要考虑时间消耗, 因为许多研究人员以牺牲过多的时间消耗为代价来提高模型的检测率。这对于入侵检测来说是不可取的, 因为大量的时间消耗会影响其实用性。为了减少分类器集成的时间消耗, 可以考虑采用数据降维方法, 如主成分分析(PCA)[13], 概率主成分分析(PPCA)[14], 以及核主成分分析(KPCA)[15]。数据降维是过滤数据噪声

的有效方法, 它可以在预处理阶段尽可能地减少数据而不影响检测结果, 从而大大降低了时间消耗。

在本文中, 使用双层异质学习器集成学习策略(Intrusion Detection Model using Double-layer Heterogeneous Ensemble Learner Strategy, IDHEL)设计了一种新颖的入侵检测模型。首先, 采用数据预处理方法来降低计算量, 以减少时间消耗。然后, 我们使用五个异质分类器对数据集执行异常检测, 采用分层十倍交叉验证。接下来, 根据分类器评估公式(Classifification Evaluation Algorithm, CEA)选择三个最佳分类器。最后, 执行基于概率加权投票的异质学习器集成算法(Multi-classififier Fusion Algorithm, McFA)。我们还将 IDHEL 模型与单一分类器(如朴素贝叶斯, Bp 神经网络, C4.5, 逻辑回归, SVM)和使用相同数据集的其他一些先进算法进行比较, 以证明我们的模型更适合入侵检测。

本文的贡献如下:

1. 提出了一种使用双层异质分类器集成学习策略(IDHEL)的新型入侵检测模型, 可以通过概率加权投票机制提高入侵检测的性能。

2. 利用概率核主成分分析方法降低了模型训练的计算成本, 并采用分层十倍交叉验证方法避免了过度拟合。

3. 在入侵检测数据集上对 IDHEL 模型进行了评估, 实验结果表明, IDHEL 模型在准确性, 错误率和时间消耗方面均具有优越性。

本文的其余部分安排如下: 第 2 节介绍了国内外相关工作; 第 3 节详细介绍了本文提出的入侵检测模型; 第 4 节介绍了实验并对结果进行了讨论; 第 5 节主要阐述了我们的工作结论。

2 国内外相关研究进展

2.1 入侵检测模型

我们将基于异常的入侵检测方法分为三类: 单一传统分类器、深度学习和集成学习。

2.1.1 单一传统分类器

单个传统分类器方法就是采用单个分类器来对数据进行训练学习。

Syarif 等人[2]讨论了五种不同的异常检测技术, 并且使用 NSL-KDD 数据集来评估网络异常检测中的聚类算法。然而, 实验结果表明采用这几种算法会产生产生较高的误报率(超过 20%)。

此外, Eslamnezhad 等人[5]设计了一种改进的 K-Means 算法, 称为 Min-Max K-Means。MinMax K-Means 是一种新的聚类算法, 试图解决 K-Means 初始化问题。该算法从随机选择簇的初始中心开始, 然后尝试应用最小化簇的最大内部方差, 而不是最小化 K-Means 算法中簇的内部方差之和。对每个聚类进行加权, 以便为具有较大内部方差的聚类分配较高权重。 通过应用此方法, 结果变得更少依赖于初始化, 并且即使未最佳地选择簇的初始中心, 聚类的质量也会增加。实验结果表明, Min-Max K-Means 算法的总检测率为 81%, 而常规 K-Means 分配较高权重。 实验结果表明, Min-Max K-Means 算法的总检测率为 81%, 而常规 K-Means 算法的总检测率为 75%。然而, 他们都只是尝试改进了 K-Means 算法, 尽管结果证明他们所提出的算法比 K-Means 算法要好, 但是他们并没有尝试和其他的算法进行比较, 也没有尝试减少时间开销。

SVM 算法由于使用核函数可以向高维空间进行映射和解决非线性的分类, 以及分类思想简单, 就是将样本与决策面的间隔最大化等优点, 也被许多研究者所采用。

Bo 等人[3]通过使用短序列数据, 使用 SVM 模型将数据标记为异常或正常。他们的基于 SVM 的模型有较高的准确率, 并且可以有效地进行实时入侵检测, 其框架图如图 2 所示。然而他们仅仅考虑了提高使用 SVM 算法的检测率和减少使用 SVM 算法的时间消耗。根据我们的实验可以得知, 使用 SVM 算法本身所耗费的时间就不多。因此在此基础上减少时间消耗的意义不是很大, 此外, 他们忽略了与其他分类算法相对比。

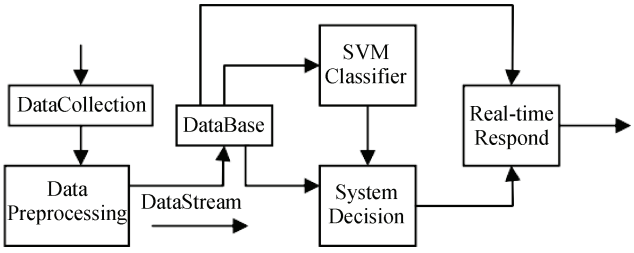

图 2 Bo 等人的提出的 SVM 框架[3]

Figure 2 Bo et al.'s proposed SVM framework[3]

2.1.2 深度学习

深度学习指含多个隐藏层的多层感知器的学习结构。深度学习通过组合低层特征以形成更加抽象的高层表示属性类别或特征, 以发现数据的分布式特征表示。

Tang 等人[4]构建了一个深度神经网络(DNN)模型, 在 SDN 环境中获取数据集的六个基本特征, 以此基于流量进行异常检测。然而, 他们所提出的模型

最后实验的准确率只有 75.75%, 并不是很高, 此外模型的可解释性不是很好。Shone 等人[11]提出了一种用于无监督特征学习的非对称深度自动编码器(NDAE), 并提出了一种使用堆叠 NDAE 的深度学习分类模型。如图 3 所示, 模型使用了堆叠排列的两个 NDAE, 并与 RF 算法结合使用。每个 NDAE 有 3 个隐藏层, 每个隐藏层使用与特征相同数量的神经元(由图中的编号表示)。通过交叉验证多种组合(即神经元和隐藏层的数量)确定这些确切的参数, 直到确定最有效。研究表明作者所提出的模型误报率较高, 且时间耗费较大。此外, 模型的可解释性较差。

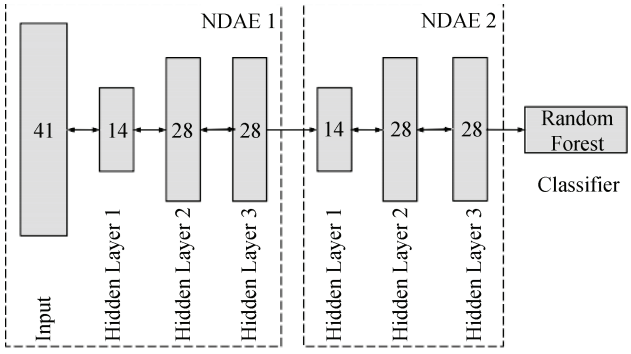

图 3 堆叠 NDAE 的深度学习分类模型[11]

Figure 3 Deep learning classification model for stacked NDAE[11]

2.1.3 集成学习

集成学习指通过合并多个分类器来提升机器学习性能, 这种方法相较于采用单个分类器的方法通常能够获得更好的预测结果。

Tengl 等人[6]提出了一种基于遗传算法(GA)的集成分类器最优加权策略的协同鲁棒入侵检测模型。在所提出的模型中, 如图 4 所示, 他们使用 PCA 来进行数据降维, GA 用于优化集合分类器的每个基本分类器的权重。然而, 他们所提出的方法虽然最后得到的准确率较高, 但却花费了大量时间来调整权重, 这将会影响算法的实时性和可用性。

此外, Sornsuwit 等人[7]采用 Adaboost 算法创建决策树、朴素贝叶斯, SVM 和 MLP 分类器的集合, 并通过实验结果证明了优越性。

2.2 数据降维算法

为了提高模型的实用性, 许多研究人员将降维方法应用到数据预处理阶段以减少时间消耗。PCA 是一种经典的数据降维算法, 一些研究人员针对上述问题改进了 PCA 算法。例如, Ge 等人[15]提出了一个可以自动确定潜变量有效维数的模型。对于具有多种运行模式的监测过程, 将贝叶斯正则化方法扩展到其混合形式, 然后开发了混合贝叶斯正则化的 PPCA 方法。

此外, Li 等人[5]提出了一种基于 KPCA 和最小二乘支持向量机(LSSVM)的非线性过程异常检测与诊断方法。这项工作试图尽可能地降低计算成本, 同时确保准确性, 以提高模型的实用性。

这些减少数据维度的尝试都是有意义的。它们都在确保准确率的情况下, 尽可能降低计算成本, 从而提高模型的实用性。

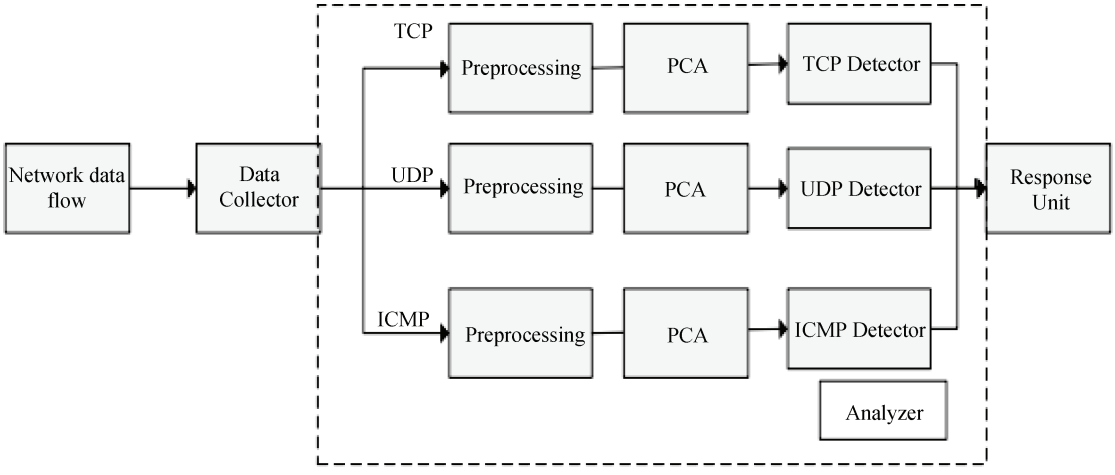

图 4 协同入侵检测的体系结构图[6]流程图

Figure 4 The architecture of collaborative intrusion detection[6]

总结相关工作, 可以发现尽管使用 K-Means 和神经网络等机器学习算法在基于异常的入侵检测中实现了较高的检测率和准确率, 但在大多数情况下都会导致较高错误率并产生巨大的计算开销, 影响其实用性。基于异常的 IDS 的高时间消耗主要归因于这些 IDS 需要分析的数据中存在大量的特征。因

此, 本文旨在通过异质学习器集成策略来解决异常检测的较高错误率的问题和通过数据降维技术解决基于异常的 IDS 的高计算成本问题。

3 入侵检测 IDHEL 模型

3.1 模型概述

我们所提出的 IDHEL 模型分为两个部分。

第一部分是数据预处理。使用 PKPCA 数据降维算法, 它结合了 PPCA 和 KPCA 的优点, 能够尽可能地减少信息损失和降低计算开销。

第二部分是入侵检测的双层策略, 又分为两部分。首先是单一分类器, 使用五种不同的分类器来分别检测, 包括: 朴素贝叶斯, Bp 神经网络, C4.5 决策树, 逻辑回归和 SVM。这是因为根据 Zaman 等人[16]和 Syarif 等人[2]的工作, 这五个分类器的分类结果相对较好且可解释较高。此外, 应用了分层十折交叉验证方法来防止模型过拟合。其次是异构的集成学习器策略, 使用分类器评估算法(CEA)选择最好的三个分类器作为组件学习器, 然后执行多分类器融合算法(McFA)进行再处理。整个模型的体系结构如图 5 所示。

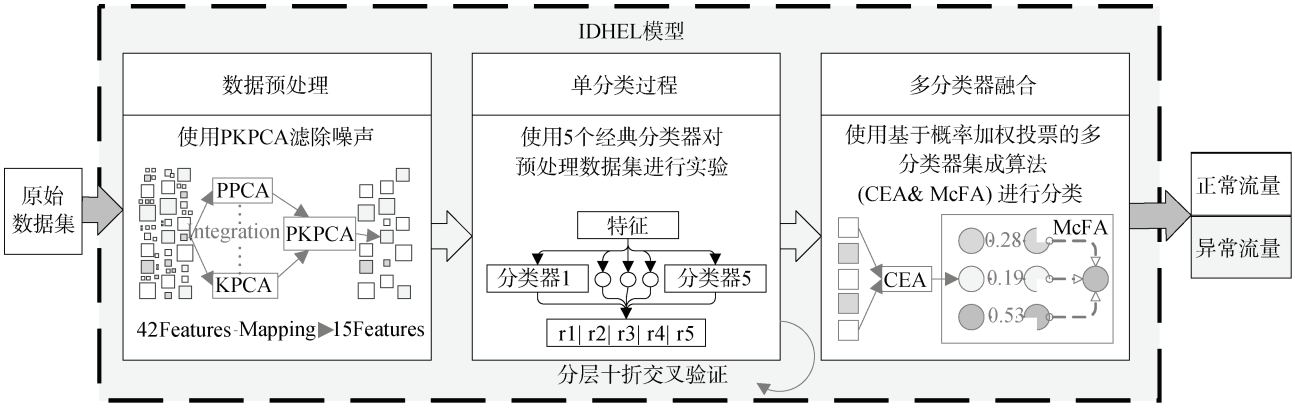

图 5 IDHEL 模型框架图

Figure 5 The overall framework of IDHEL model.

3.2 数据降维算法

根据 Sornsuwit 等人[7]的描述, 减少特征能够提高入侵检测中弱学习器的分类效率。数据维度本质上是从一个维度空间映射到另一个维度空间, 特征的个数并没有减少, 然而在映射的过程中特征值会发生相应的变化。

PCA 是一种线性投影技术, 遵循最大化数据方差的原则来进行数据降维, 尽可能地保留有效信息。使用 PCA 的降维过程首先需要对数据集的特征值进行归一化处理, 接着求协方差的特征值和特征向量, 特征向量都归一化为单位向量, 然后将特征值按照从大到小的顺序排序, 选择其中最大的 k 个, 接下来将其对应的 k 个特征向量分别作为列向量组成特征向量矩阵, 最后, 将样本点投影到选取的特征向量上。

然而, PCA 算法存在以下两个问题。首先, PCA 没有将数据的概率分布考虑; 其次, PCA 仅考虑了数据的二阶统计信息, 而没有利用高阶统计信息, 忽略了数据的非线性相关性。

针对上述两个问题, 前人分别对 PCA 进行了改进。

PPCA 对 PCA 做了概率上的解释, 延伸了 PCA 算法。它是一种考虑每个变量概率分布的方法, 在确定主元和误差的概率函数后, 通过期望最大(EM)算法建立模型。其具体步骤如下:

1. 将原始数据按列组成 n 行 m 列矩阵 $\boldsymbol{X}$;

2. 将原始训练样本数据进行标准中心化处理得到 $\boldsymbol{X}$;

3. 在隐含变量 x 的条件下得到观测数据的概率分布;

4. 采用 EM 算法获得概率 PCA 的模型参数 W (因子矩阵)和其方差;

5. 删除不满足因子矩阵与方差特定关系的归一化数据;

6. 剩余满足条件的数据即为降维到 k 维后的数据。

核主成分分析(KPCA)则通过非线性变换将数据映射到高维, 并提取高维空间中的特征以改进特征提取。其具体步骤如下所示:

1. 将原始数据按列组成 n 行 m 列矩阵 $\boldsymbol{X}$;

2. 计算核矩阵, 选定高斯径向核函数中的参数, 计算核矩阵 $\boldsymbol{K}$, 修正核矩阵得到 KL;

3. 求出协方差矩阵 $\boldsymbol{C}$, 运用 Jacobi 迭代算法计算 KL 的特征值和特征向量;

4. 将特征向量按对应特征值大小从上到下按行排列成矩阵, 取前 k 行组成矩阵;

5. 通过施密特正交化方法单位正交化特征向量得到 P;

6. $Y=PX$ 即为降维到 k 维后的数据。

PPCA 和 KPCA 分别改进了 PCA 存在的两个问题, 因此, 我们可以考虑将两种算法结合起来, 既能够将数据的概率分布考虑进去, 又能够利用数据的高阶统计信息, 以此来得到更好的降维效率。

因此, 在本文中, 我们使用概率核主成分分析(PKPCA)方法, 该方法不仅能够捕获数据的高维信息, 而且还考虑了其概率分布[17]。该方法具体描述如下:

假设 $\{x_1,x_2,...,x_N\}$ 是数据空间 R^d 中的训练数据, 并且数据由映射函数 Ψ 被映射到高维数据空间 R^f 中, 其中 $f>d$。映射数据用 $\Psi_{f\times N}=\{\Psi_1,\Psi_2,...,\Psi_N\}$ 表示。

隐藏变量模型是 $\boldsymbol{\Psi}(x)=\boldsymbol{W}z+\mu+\varepsilon$, 其中 $z\sim N(0,I_q)$, $\varepsilon\sim N(0,\rho I_f)$, $\boldsymbol{W}$ 表示 $f\times q$ 的因子矩阵。

根据 Tipping 等人的描述[18], 参数 μ 和 $\boldsymbol{W}$ 的最大似然估计表示为:

$$\mu=\bar{\Pi}_0=\frac{1}{N}\sum_{n=1}^{N}\Pi(x_n)=\Psi S$$

$$\boldsymbol{W}=U_q(\lambda_q-\rho I_q)^{\frac{1}{2}}R$$

其中 $\boldsymbol{R}$ 是任何 $q\times q$ 大小的旋转矩阵, λ_q 和 U_q 分别是第 q 大特征值和相应的包含 C 的特征向量。

Scholkopf 等人[19]提出了一种 EM 算法, 用于在 PKPCA 中查找参数 Q 和 $\tilde{\rho}$, 使用以下迭代公式:

$$\tilde{Q}=KQ(\rho I_q+M^{-1}Q^TK^2Q)^{-1}$$

$$\tilde{\rho}=\frac{1}{f}tr(K-KQM^{-1}Q^TK)$$

其中, $M=\rho I_q+W^TW=\rho I_q+Q^TKQ$, $\tilde{Q}$ 和 $\tilde{\rho}$ 是更新后的估算值。

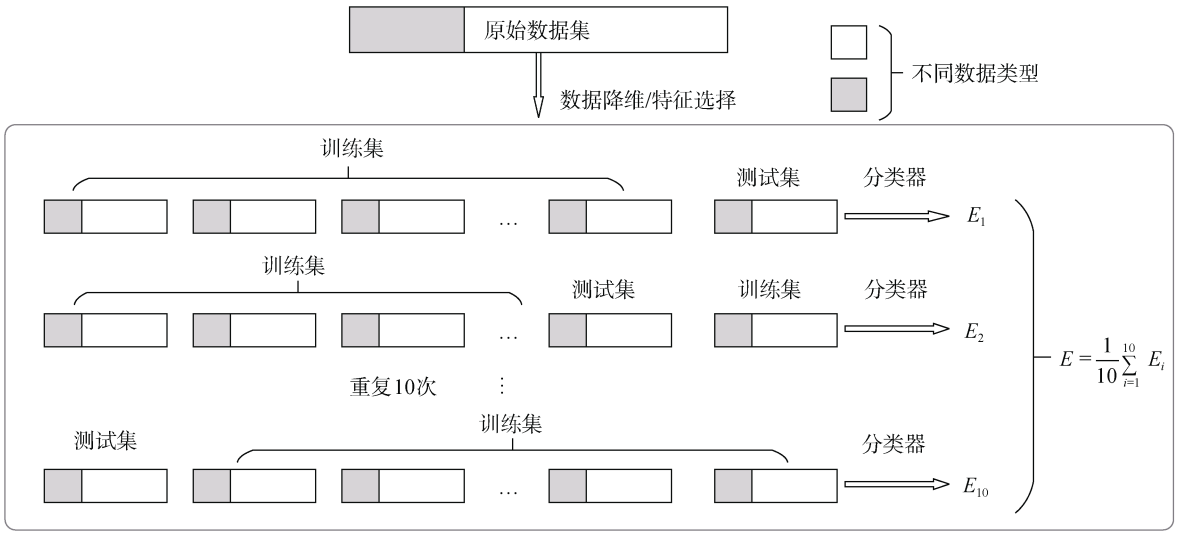

图 6 数据分层 10 折交叉验证示意图

Figure 6 Layered 10-fold cross-validation diagram of dataset

3.3 验证策略

验证策略可以评估模型的预测性能并防止过拟合。因为在现实世界中, 数据集并非全部平衡。 对于不平衡的数据集, 简单的交叉验证不考虑原始数据集的分布。本文使用分层 10 折交叉验证方法。分层意味着原始数据中每个类别的比例关系在每个折叠中保持不变。具体方法如图 6 所示。假设有两种类型的原始数据, 比例为 1∶2, 那么十个折叠中的每一个中的数据类别保持 1∶2 的比率, 这使得结果更可靠。

采用分层 10 折交叉验证算法来进行分类的学习器示意图如图 7 所示。首先将数据集按类别等比例划分成 10 份, 用 9 份作训练集, 1 份作测试集, 每个分类器经过 10 次交叉验证, 最终得到五种分类器结果。接着, 进入下一步基于概率投票加权的分类器集成算法。

3.4 基于概率加权投票的异质学习器集成算法

相关工作表明[2-3,5], 在分类过程中, 单个分类器可能带来分类偏差, 导致模型具有较高的错误率。集成策略意味着融合多个分类器可以产生更好的结

果。如果我们以合理的方式融合多个异质学习器, 我们可能就会得到理想的分类结果, 并且整体分类误差也会减少。因此, 我们使用双层异质学习器集成策略来进行入侵检测。在 4.3 节中, 详细介绍了分别使用五种不同的分类器和我们在本文中提出的 IDHEL 模型进行实践的差别。

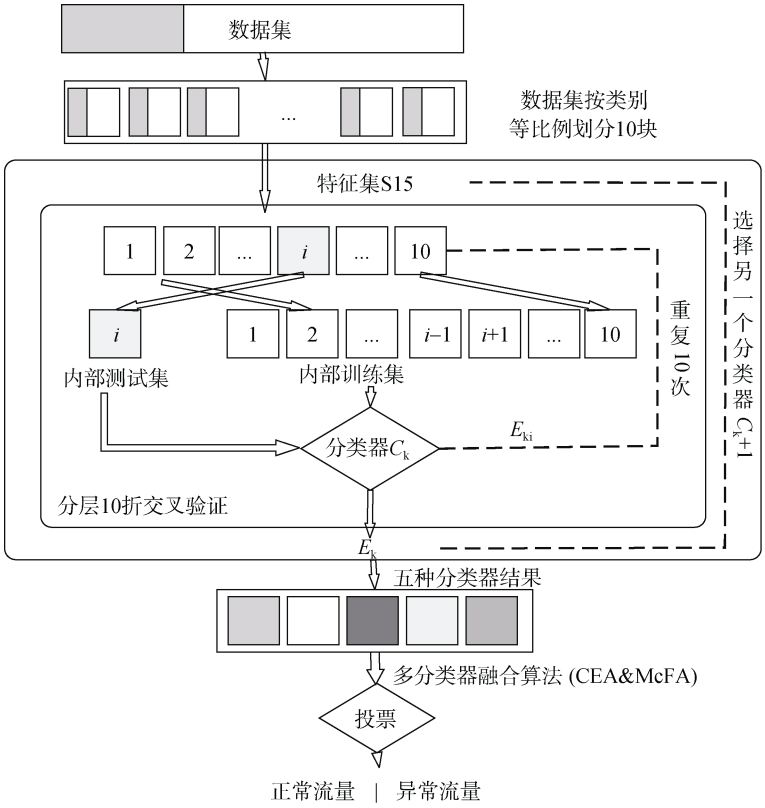

图 7　分类器分层 10 折交叉验证算法过程

Figure 7　The process of learners using the layered 10- fold cross-validation algorithm for classification

在本文中, 为了显著提高实验效果, 我们采用基于概率加权投票的异质学习器集成算法来进行入侵检测。该算法主要包括分类评估和多分类器集成两个步骤。

3.4.1　分类评估算法(CEA)

我们使用以下公式来评估每个分类器的效果:

$$\gamma_{CEA}=\frac{2\times F\times AUC}{F+AUC}$$

其中, F 表示 F-Measure, AUC 表示 AUC 的值, 即 ROC 曲线下的面积。

根据 CEA 公式, 我们可以选择出针对该数据集适用于 PKPCA 的三种效率最高的分类算法。

尽管存在许多分类指标, 例如: recision, recall, F-Measure, ROC, AUC 等, 然而单独的高精确率和高召回率并不能够证明该算法是有效的, 而 F 值则是对精度和召回率的综合评估, 它是两者的调和平均值, 如下公式所示:

$$F=\frac{2\times precision\times recall}{precision+recall}$$

ROC 曲线则将 false positive rate(FPR)作为横坐标, true positive rate 作为纵坐标, 它能够很容易地查出任意界限值时的对性能的识别能力, 我们可以通过分别计算各个实验的 ROC 曲线下的面积(AUC)来比较实验结果的优劣。

$$AUC=1-\frac{1}{m^+m^-}\sum_{x^+\in D^+}\sum_{x^-\in D^-}(W(f(x^+)<f(x^-))+\frac{1}{2}W(f(x^+)=f(x^-)))$$

其中, 正样本个数为 m^+, 负样本个数为 m^-, D^+ 为所有正例组成的集合, x^+ 是其中的一个正例, D^- 为所有反例组成的集合, x^- 是其中的一个反例, $f(x)$ 是模型对样本 x 的预测结果, 在 0—1 之间, W 仅在 x 为真时取 1, 否则取 0。

由上分析可以看出, F 值和 AUC 能够比较直观地评判分类效果, 因此我们使用了 F-Measure 和 AUC 的调和平均数来综合评判分类效果。

3.4.2　多分类器集成算法(McFA)

本文采用了概率加权投票的方式来集成多个分

类器。投票法是最简单也是最广泛的集成方法, 这种方法是对各个分类器的判决进行投票, 其最大得票的判决作为最后系统的识别结果。

假设给定的模式空间由两个互斥的集合构成, 即 $S = D_1 \cup D_2$, 若分类器 c_i 对于来自 D_j 的样本有一个期待的输出向量 $\boldsymbol{E}_i^j$, 并且有 $\boldsymbol{E}_i^1 = (0,1)$, $\boldsymbol{E}_i^2 = (1,0)$。那么, 当分类器 c_i 有一个输出向量 $\boldsymbol{y}_i = (0.9, 0.1)$ 或 $\boldsymbol{y}_i = (0.6, 0.4)$ 时, 分类器都会将这两个输入向量识别为相同的类 D_1。然而, 对于输出向量 $\boldsymbol{y}_i = (0.6, 0.4)$ 来说, 它的分类效果显然比不上 $\boldsymbol{y}_i = (0.9, 0.1)$。因此, 可以考虑给概率 $P\{S \in D_j \mid c_i(x_k) = \boldsymbol{y}_i\}$ 赋予不同的权重值。

由于本文所采用的三种分类算法的输出向量并不一致, 所以我们在进行多分类器融合之前, 首先要将输出结果转换成统一的概率模式, 然后再计算各个分类器的加权值。当满足 $c_i(x_i) = \boldsymbol{y}_i$ 时, 对于各个分类器 c_i 的概率加权定义为:

$$P\{S \in D_j \mid c_i(x_k) = \boldsymbol{y}_i\} = \frac{\exp(-\left|\boldsymbol{E}_i^j - \boldsymbol{y}_i\right|)}{\sum_{j=1}^{M} \exp(-\left|\boldsymbol{E}_i^j - \boldsymbol{y}_i\right|)}$$

其中 $\boldsymbol{E}_i^j$ 是输入类别为 D_j 时分类器 c_i 的期望输出。将 $P\{S \in D_j \mid c_i(x_i) = \boldsymbol{y}_i\}$ 作为投票表决时分类器 c_i 的第 j 个输出的得票数目, 则 $S \in D_j$ 的总得票数为:

$$P(S \in D_j) = \sum_{i=1}^{I} p(S \in D_j \mid c_i(x_i = y_i)),\ j = 1, 2, \cdots, M$$

因此, 基于异质学习器的输出向量加权投票表决规则表示为:

$$E(S) = \begin{cases} k,\ \textit{if } k \in \Lambda, \textit{and } P\{S \in D_k\} = j \in \\ \{\Lambda P\{S \in D_j\} \geqslant T_k \\ 0,\ \textit{otherwise refuse to identify} \end{cases}$$

其中 T_k 是表决阈值, 它能根据不同的应用需求设定不同的值。此外, 为了提高算法的可靠性, 本文采用了拒绝识别的方法, 其时间复杂度为 $O(n)$。

4 实验

4.1 数据集说明

我们使用 NSL-KDD 数据①, 它是开源 KDD99 [20] 的修改版本。与 KDD 数据集相比, NSL-KDD 数据集具有以下几个优点: (i)没有冗余记录, (ii)没有重复记录, (iii)训练和测试中的记录数量设置合理。因此, 不同研究工作的评价结果将是一致的和可比的。

在 NSL-KDD 数据集中, 总共有 148 517 个数据, 77 054 个正常数据和 71 463 个异常数据。

4.2 实验设置

我们将提出的 IDHEL 模型与五个单独使用的分类器进行比较, 这些分类器是朴素贝叶斯, Bp 神经网络, C4.5, 逻辑回归和 SVM, 以及其他在 NSL-KDD 数据集上进行实验的入侵检测模型。

- MinMax K-means(Eslamnezhad 等人, 2015): 该算法克服了 K-means 算法中对初始中心的敏感性不足的问题[21]。
- 改进的 K-means 算法(Wang, 2011): 通过尽可能选择初始中心来克服初始中心选择的灵敏度问题[22]。
- 改进的 SVM 算法(Heba et al. 2013): 该算法基于主成分分析(PCA)和支持向量机(SVM)[14]。
- DNN(Tang et al. 2016): 在该模型中, 建立了深度神经网络(DNN)模型, 该模型在 SDN 环境中获得了六个基本特征[10]。
- S-NADE(Tang et al. 2018): 它提出了一种用于特征学习的非对称深度自动编码器(NDAE)和一种使用堆叠 NDAE 的新的深度学习分类模型[11]。
- Ensemble with weight strateg(Tengl et al. 2018): 在这个模型中, 采用遗传算法(GA)来优化每个基本分类器的权重, 采用 PCA 来进行数据降维[6]。
- Adaboost Ensemble(Sornsuwit et al. 2016): 在这个模型中, 采用了 Adaboost 算法创建弱学习者的集合, 以提高分类器的性能[7]。

4.3 数据预处理实验结果

为了证明 PKPCA 的必要性和优越性, 我们进行了三个独立的实验: 直接使用五个分类器而不进行数据预处理, 和使用 PCA 和 PKPCA 数据降维之后再进行分类。表 2 显示了这三个实验中使用的特征数, 分类的准确性和时间消耗。

从表中可以看出, 将 41 个特征降到 15 个特征之后, 使用朴素贝叶斯分类器的实验结果的准确性没有降低。此外通过 PKPCA 进行数据预处理之后, 使用 BP 神经网络的结果增加到 97.07。由此可见, 减少数据维度不会对分类的准确性产生过度的负面影

① https://www.unb.ca/cic/datasets/nsl.html

响。相反, 在数据预处理之后, 五个分类器的运行时间都有了显著的下降。PCA 的平均时间消耗降低了 74.8%, PKPCA 的平均时间消耗降低了 71.1%。

4.4 入侵检测性能比较

在异质学习器集成的部分中, 我们使用了基于概率加权投票的多分类器集成算法。 图 5 显示了经过 PKPCA 数据降维处理之后, 采用五种不同学习器进行分类的精度、AUC 值和 F 值。我们使用 CEA 公式来选择三个最佳分类器, 图 9 显示了每个分类器的 γ_{CEA} 值。

从图 8、图 9 中可以看出, 逻辑回归、C4.5 和 SVM 这三个分类器在此数据集上具有最佳的分类效果, 因此我们选择这三个分类器使用 McFA 算法进行多分类器融合以进一步提高效率。

在使用基于概率加权投票的 McFA 算法之后, IDHEL 模型的准确率、精确率、错误率、F 值和 AUC 值的能力表图如图 9 所示。

通过图 8~图 10 可以看出, 在使用基于概率加权投票的分类器集成算法之后, 实验效果得到了显著改善。在进行异质学习器集成之前, 经过 PKPCA 数据降维预处理之后的数据集, 在单独分类器上所获得的 F-Measure 和 AUC 的最高值分别为 0.976 和 0.988, 这两者都是使用 C4.5 分类算法获得的。在使用本文提出的基于概率加权投票的分类器集成算法后, F-Measure 和 AUC 都提高到了 0.985 和 0.992。准确率为 0.957, 精确率为 0.962, 高于现今大部分主流模型。此外, IDHEL 模型在错误率方面也表现良好, 错误率小于 10%。

表 2 五种分类器在原始数据集、经过 PCA 处理的数据集和 PKPCA 处理的数据集上的分类结果比较

Table 2 Comparison of the results of classification of original dataset, PCA-processed dataset, and PKPCA-processed dataset

		分类器				
		朴素贝叶斯	Bp 神经网络	C4.5	逻辑回归	SVM
原始数据集	特征数量	41	41	41	41	41
	准确率 Accuracy(%)	80.57	97.05	98.50	91.46	96.62
	时间消耗(s)	2	1595	26	50	220
经过 PCA 数据降维处理后的数据集	特征数量	15	15	15	15	15
	准确率 Accuracy(%)	**84.20** ↑	93.62	97.09	87.93	94.05
	时间消耗(s)	**1** ↓	**326** ↓	**6** ↓	**11** ↓	**133** ↓
经过 PKPCA 数据降维处理后的数据集	特征数量	15	15	15	15	15
	准确率 Accuracy(%)	**84.73** ↑	**97.07** ↑	89.92	89.92	96.29
	时间消耗(s)	**1** ↓	**357** ↓	**7** ↓	**14** ↓	**168** ↓

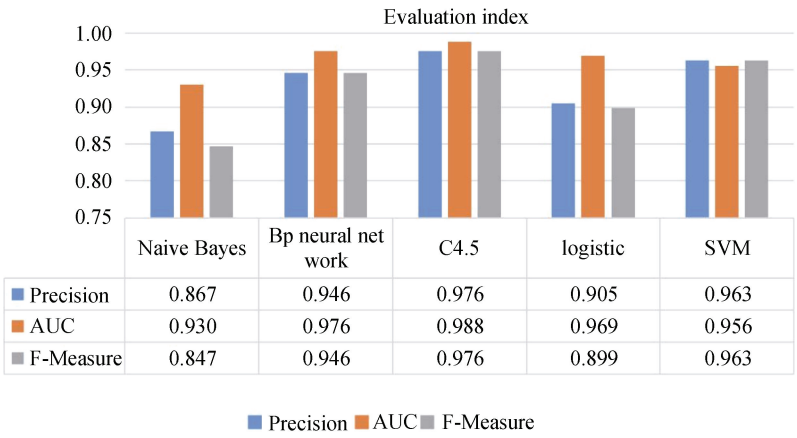

图 8 执行 PKPCA 后五种不同分类器的评估指数

Figure 8 Evaluation index for five different classifiers after performing PKPCA

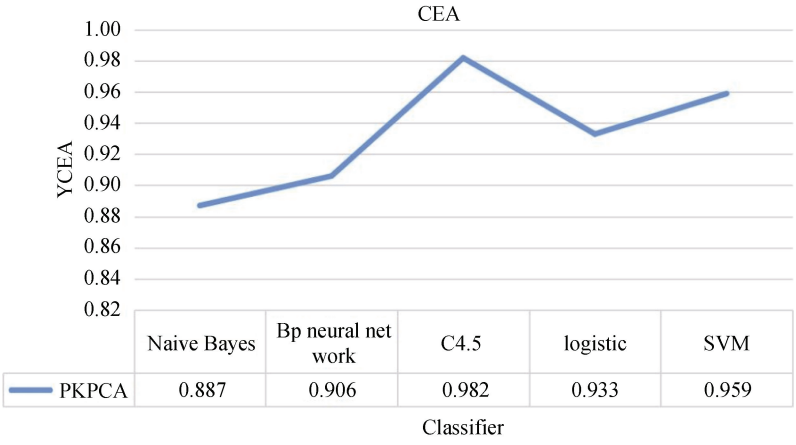

图 9　执行 PKPCA 后，五种不同分类器的 γ_{CEA} 值

Figure 9　CEA values (γ_{CEA}) for five different classifiers after performing PKPCA

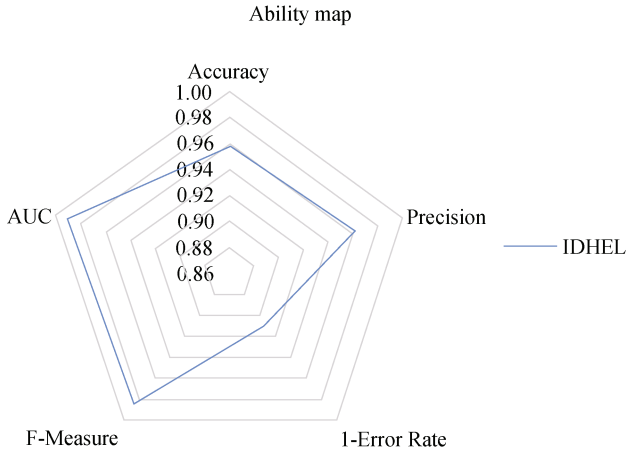

图 10　IDHEL 模型的能力表图

Figure 10　Ability map of IDHEL

4.5　时间消耗比对分析

在现实世界中，对入侵检测的实时性要求并不低。因此，我们还比较了算法的整个运行时间，如图 11 所示。从图中可以看出，本文提出的 IDHEL 模型的最终总时间消耗远小于 Bp 神经网络。在保持高准确率度和较低错误率的情况下，完整的入侵检测时间仍然小于 500s。由于也可证明，采用本文所提方法能够有效减少分类器集成的时间消耗，提高入侵检测的实时性。

4.6　模型比较

为了更好地评估 IDHEL 模型的性能，我们将其与在同一数据集上进行实验的多种主流模型进行了实验效果比较。从表 3 中可以看出我们的算法在 True Positive rate(TP)、False Positive rate(FP)、准确率和时间消耗这四个指标方面都具有优越性。

IDEHL 算法的 TP 和 FP 值在这些算法中表现最佳，分别为 0.989 和 0.061。另外，它的准确度也很高，为 0.957。

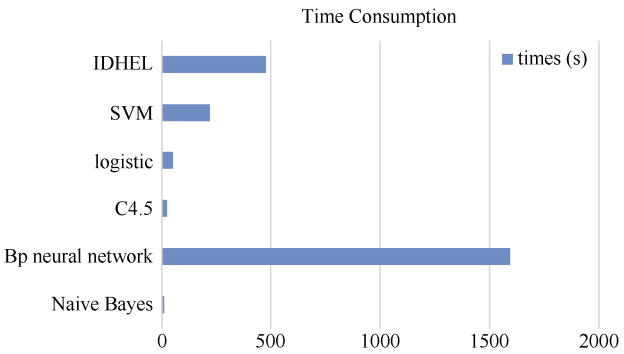

图 11　IDHEL 模型与单独分类器时间消耗方面的比较

Figure 11　Comparison of time consumption between IDHEL model and five separate classifiers

表 3　与多种主流模型的综合对比实验

Table 3　Intrusion detection algorithm comparison

类别	算法	TP	FP	准确率	时间消耗(s)
单个分类器	*MinMax K-means*	0.8136	0.094	-	-
	改进的 *K-means Algorithm*	0.68	0.43	-	492
	改进的 *SVM Algorithm*	0.796	-	0.816	-
深度学习	*DNN*	-	-	0.758	893
	S-NADE	0.765	-	0.892	722.54
集成学习	*Ensemble with weight strategy*	-	0.015	0.872	4098
	Adaboost Ensemble	-	-	0.985	1331
	IDHEL	**0.989**	**0.061**	**0.957**	479

5 结论

在本文中, 我们采用双层异质学习器的集成策略提出了一种新颖的入侵检测模型。我们使用 PKPCA 数据降维算法来来解决基于异常的 IDS 的高计算开销的问题。接下来, 我们使用多个异质学习器和分层十折交叉验证策略来执行异常检测, 并通过 CEA 公式选择在该数据集上表现最好的三个分类器。然后, 我们提出了一种基于概率加权投票的集成算法, 以进一步增强实验结果。通过实验, 我们证明了 IDHEL 模型可以实现较高精确率、较低错误率、较少时间消耗的目标。

其实, 在入侵检测方面, 本实验仍然有可改进之处。比如, 可以将程序放到 spark 架构上进行分布式处理。不过, 由于 NSL-KDD 数据集的数据量不够大, 使用分布式处理的方式, 数据分发的时间会大大超过数据处理的时间, 因此, 本文并没有采用分布式的方法进一步减少时间消耗。但是, 在现实世界中, 分布式的方法还是可取的, 它能够在 IDHEL 模型的基础上, 进一步大幅度减少时间消耗。

致 谢 这项工作得到了中国自然科学基金(No. 61702508, No.61802404), 国家重点研发计划课题(2016YFF0204002, 2016YFF0204003), “十三五”装备预研领域基金(6140002020115)的支持。这项工作也得到了中国科学院网络评估技术重点实验室和北京市网络安全与保护技术重点实验室的部分支持。

参考文献

[1] Viegas E, Santin A, Abreu V, et al. Stream Learning and Anomaly-Based Intrusion Detection in the Adversarial Settings[C]. *2017 IEEE Symposium on Computers and Communications*, 2017: 773-778.

[2] Syarif I, Prugel-Bennett A, Wills G. Unsupervised Clustering Approach for Network Anomaly Detection[M]. *Networked Digital Technologies*. 2012.

[3] Bo L, Yuan C Y. The Research of Intrusion Detection Based on Support Vector Machine[C]. *2009 International Conference on Computer and Communications Security*, 2009: 21-23.

[4] Tang T A, Mhamdi L, McLernon D, et al. Deep Learning Approach for Network Intrusion Detection in Software Defined Networking[C]. *2016 International Conference on Wireless Networks and Mobile Communications*, 2016: 258-263.

[5] Li Z H, Zhang Y, Jiang L Y. Fault Detection and Diagnosis Based on KPCA-LSSVM Model[J]. *2009 International Conference on Measuring Technology and Mechatronics Automation*, 2009, 1: 634-638.

[6] Tengl S, Zhang Z H, Teng L Y, et al. A Collaborative Intrusion Detection Model Using a Novel Optimal Weight Strategy Based on Genetic Algorithm for Ensemble Classifier[J]. *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*, 2018: 761-766.

[7] Sornsuwit P, Jaiyen S. Intrusion Detection Model Based on Ensemble Learning for U2R and R2L Attacks[C]. *2015 7th International Conference on Information Technology and Electrical Engineering*, 2015: 354-359.

[8] Wang S H. Research of Intrusion Detection Based on an Improved K-Means Algorithm[C]. *2011 Second International Conference on Innovations in Bio-inspired Computing and Applications*, 2011: 274-276.

[9] Eslamnezhad M, Varjani A Y. Intrusion Detection Based on Min-Max K-Means Clustering[C]. *7'th International Symposium on Telecommunications*, 2014: 804-808.

[10] Tang T A, Mhamdi L, McLernon D, et al. Deep Learning Approach for Network Intrusion Detection in Software Defined Networking[C]. *2016 International Conference on Wireless Networks and Mobile Communications*, 2016: 258-263.

[11] Shone N, Ngoc T N, Phai V D, et al. A Deep Learning Approach to Network Intrusion Detection[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, 2(1): 41-50.

[12] Boero L, Cello M, Marchese M, et al. Statistical Fingerprint-Based Intrusion Detection System (SF-IDS)[J]. *International Journal of Communication Systems*, 2017, 30(10): e3225.

[13] Javaid A, Niyaz Q, Sun W Q, et al. A Deep Learning Approach for Network Intrusion Detection System[C]. *BICT'15: Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies*. 2016: 21-26.

[14] Heba F E, Darwish A, Hassanien A E, et al. Principle Components Analysis and Support Vector Machine Based Intrusion Detection System[C]. *2010 10th International Conference on Intelligent Systems Design and Applications*, 2010: 363-367.

[15] Ge Z Q, Song Z H. Mixture Bayesian Regularization Method of PPCA for Multimode Process Monitoring[J]. *AIChE Journal*, 2010, 56(11): 2838-2849.

[16] Zaman M, Lung C H. Evaluation of Machine Learning Techniques for Network Intrusion Detection[C]. *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018: 1-5.

[17] Xie Y C, Wang H Q, Li P. PKPCA: A Nonlinear Principal Component Analysis Algorithm Integrating Priori Class Information[J]. *Journal of Circuits and Systems*, 2003, 8(6): 95-99, 81.
(解应春, 王海清, 李平. PKPCA: 融合先验类别信息的非线性主元分析算法[J]. *电路与系统学报*, 2003, 8(6): 95-99,81.)

[18] Tipping M E, Bishop C M. Mixtures of Probabilistic Principal Component Analyzers[J]. *Neural Computation*, 1999, 11(2): 443-482.

[19] Schölkopf B, Smola A, Müller K R. Kernel Principal Component Analysis[M]. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997: 583-588.

[20] KumarShrivas A, Kumar Dewangan A. An Ensemble Model for Classification of Attacks with Feature Selection Based on KDD99 and NSL-KDD Data Set[J]. *International Journal of Computer Applications*, 2014, 99(15): 8-13.

[21] Eslamnezhad M, Varjani A Y. Intrusion Detection Based on Min-Max K-Means Clustering[C]. *7'th International Symposium on Telecommunications*, 2014: 804-808.

[22] Wang S H. Research of Intrusion Detection Based on an Improved K-Means Algorithm[C]. *2011 Second International Conference on Innovations in Bio-inspired Computing and Applications*, 2011: 274-276.

[23] Effendy D A, Kusrini K, Sudarmawan S. Classification of Intrusion Detection System (IDS) Based on Computer Network[C]. *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering*, 2017: 90-94.

[24] Shao M, Kim M S, Valgenti V C, et al. Grammar-Driven Workload Generation for Efficient Evaluation of Signature-Based Network Intrusion Detection Systems[J]. *IEICE Transactions on Information and Systems*, 2016, E99.D(8): 2090-2099.

[25] Kreimel P, Eigner O, Tavolato P. Anomaly-Based Detection and Classification of Attacks in Cyber-Physical Systems[C]. *International Conference on Availability*. 2017.

[26] Zhong J, Deng X B, Wen L S, et al. An Unsupervised Network Intrusion Detection Based on Anomaly Analysis[C]. *2009 Second International Conference on Intelligent Computation Technology and Automation*, 2009: 367-370.

[27] Alrawashdeh K, Purdy C. Toward an Online Anomaly Intrusion Detection System Based on Deep Learning[C]. *2016 15th IEEE International Conference on Machine Learning and Applications*, 2016: 195-200.

[28] Hajisalem V, Babaie S. A Hybrid Intrusion Detection System Based on ABC-AFS Algorithm for Misuse and Anomaly Detection[J]. *Computer Networks*, 2018, 136: 37-50.

[29] Jose S, Malathi D, Reddy B, et al. A Survey on Anomaly Based Host Intrusion Detection System[J]. *Journal of Physics: Conference Series*, 2018, 1000: 012049.

[30] Van N T, Bao H, Thinh T N. An Anomaly-based Intrusion Detection Architecture Integrated on OpenFlow Switch[C]. *International Conference on Communication & Network Security*. 2016.

凌玥 于 2017 年在南京邮电大学信息安全专业获得学士学位。现在中国科学院大学网络空间安全专业攻读硕士学位。研究领域为网络安全态势感知、入侵检测等。Email: lingyue@iie.ac.cn

刘玉岭 于 2013 年于中国科学院大学获得博士学位。现任中国科学院信息工程所高级工程师, 硕士生导师, 研究方向为网络安全态势感知、网安大数据分析。Email: openingliu@126. com

姜波 于 2016 年在中国科学院大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所助理研究员。研究领域为网络安全态势感知、知识图谱、数据挖掘等。Email: jiangbo@iie.ac.cn

李宁 于 2014 年在中国科学院计算技术研究所计算机软件与理论专业获博士学位。现任中国科学院信息工程研究所助理研究员。研究领域为网络安全态势感知, 网络威胁信息验证、数据挖掘等。Email:lining6@iie.ac.cn

卢志刚 于 2010 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所高级工程师, 中国科学院网络空间安全学院副教授。研究领域为网络安全态势感知、网络攻击检测、移动终端安全等。
Email: luzhigang@iie.ac.cn

刘宝旭 于 2002 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所研究员, 第六研究室主任。研究领域为网络安全攻防对抗、网络安全测评技术等。
Email: liubaoxu@iie.ac.cn