计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200500001

基于 GR-AD-KNN 算法的 IPv6 网络 DoS 入侵检测技术研究

赵志强　易秀双　李　婕　王兴伟
东北大学计算机科学与工程学院　沈阳 110819
(zhaozq0518@foxmail.com)

摘　要　随着 IPv6 网络流量的快速增加和复杂化,传统入侵检测系统 Snort 是基于具体规则对 DoS 攻击进行检测的,这降低了 IDS 的检测性能。为了解决 IPv6 网络环境下的 DoS 入侵检测问题,采用了机器学习中的轻量级 KNN 的优化算法。首先,通过信息增益率实现特征的双重降维,针对具有较多类型子特征的离散特征进行选择和聚合,以实现进一步降维,减小实际运算的特征维度。其次,利用信息增益率作为优化样本欧氏距离测量的权重。基于所提出的反向距离影响力的度量指标,对 KNN 算法的分类决策算法进行了优化,使检测技术的效果得到进一步提高。实验结果表明,相比传统基于平均距离的 TAD-KNN 算法和仅优化距离定义的 GR-KNN 算法,GR-AD-KNN 算法在 IPv6 网络流量特征检测中不仅可以提升整体检测性能,同时还对小群体样本分类拥有更好的检测效果。

关键词:GR-AD-KNN 算法;平均增量距离决策;双重降维;IPv6;信息增益率
中图法分类号　TP393.0

Research on DoS Intrusion Detection Technology of IPv6 Network Based on GR-AD-KNN Algorithm

ZHAO Zhi-qiang, YI Xiu-shuang, LI Jie and WANG Xing-wei
College of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

Abstract　With IPv6 network traffic rapidly increasing, the traditional intrusion detection systems, such as Snort, based on specific rules to detect DoS intrusion attacks, have the poor performance and adaptability in detecting DoS attacks. In order to solve the problem of detecting DoS attacks in IPv6, the KNN algorithm is improved in this paper. First, in order to decrease the number of low influential sub-features of discrete type features, the approach of selecting and clustering of sub-feature is implemented by information gain ratio, which can decrease the number of features and improve the efficiency in detecting DoS attack in IPv6. Second, the improved algorithm GR-AD-KNN using information gain ratio as the weight of features to change Euclidean distance is proposed to achieve DoS attack detection. Based on a metric about reverse distance influence, the classification decision method in KNN algorithm is optimized, then the accuracy of detection approach is further improved. Experiments show that, compared with the TAD-KNN algorithm based on the average distances to classify attacks and the GR-KNN algorithm which only optimizes the Euclidean distance definition, the GR-AD-KNN algorithm not only improves the overall detection performance in IPv6 network traffic features detection, but also has better detection results on small population attack samples.

Keywords　GR-AD-KNN algorithm, Average increment distance classification, Twice reducing dimensionality of features, IPv6, Information gain ratio

1　引言

随着计算机网络的快速发展,传统的 IPv4 网络地址已逐渐耗尽。为了缓解这一问题,NAT 技术的诞生延缓了 IPv4 地址耗尽的速度,但仍然没有从根本上解决地址不足的问题。下一代互联网网络协议 IPv6 的产生,从根本上解决了这一问题,使得可以用的 IP 地址数量大幅度增加,为物联网等网络提供了大量的 IP 地址。

IPv6 与 IPv4 的不同之处,不仅在于在可用地址数量上有着较大的差异,两个版本的协议的 IP 数据报首部也有所不同。与此同时,在 IPv6 网络中取消了 ARP 等协议,取而代之的是 NDP 协议。因此,在一定程度上,两个版本的 IP 协议还是存在差别的。在安全方面,随着 IPv6 技术的快速发展,与之相关的 DoS 攻击也逐渐越来越多,IPv6 的出现也并没有根本杜绝 DoS 攻击。

针对网络中的 DoS 攻击,这类网络安全隐患可以通过网络入侵检测系统(Intrusion Detection System,IDS)进行检测。IDS 可以发现网络中传输数据的变化情况,从而及时发现网

基金项目:国家重点研发项目(2017YFB0801701);国家自然科学基金资助项目(61572123);辽宁省高校创新团队支持计划资助项目(LT2016007);赛尔网络创新项目(NGII20160616)
This work was supported by the National Key Research and Development Project(2017YFB0801701), National Natural Science Foundation of China(61572123), Program for Liaoning Innovative Research Team in University(LT2016007) and CERNET Innovation Project(NGII20160616).
通信作者:易秀双(xsyi@mail.neu.edu.cn)

络所存在的异常情况。但是，目前在网络流量数量和类型快速增加，基于具体特定规则的 IDS 在检测 IPv6 网络中的 DoS 攻击时，可能存在适应性较差和规则匹配时间较长等问题。

针对 IPv6 网络中的 DoS 攻击，本文对入侵检测技术进行了研究，实现了基于 IPv6 网络的 DoS 入侵检测的功能，具体贡献如下：

1）由于 IPv6 流量快速增加，为了提高检测效率，基于信息增益率，本文对网络流量特征进行降维，针对离散特征实现了“双重降维”的效果。在实现特征选择的基础上，又对离散类型特征进行了部分聚合，从而可以进一步减小实际运算中的特征维度，提高 IPv6 网络中的 DoS 攻击的检测效率。

2）利用信息增益率作为特征的权重，采用偏移增量平均距离的思想对 KNN 算法的决策方法进行优化，以提高 KNN 算法的稳定性。将 GR-AD-KNN（Information Gain Ratio-Average Distance-KNN）算法用于 IPv6 的 DoS 入侵检测，优化了检测性能，并减小了“小群体分类劣势”所带来的负面影响。

2 相关研究

2.1 IPv6 的 DoS 攻击

随着网络规模的快速发展，接入互联网的设备数量日益暴增，原先的 IPv4 网络地址逐渐枯竭[1]。因此，IPv6 网络的诞生解决了地址不足这一问题[2]，也为物联网和其他需要大量网络地址的产业的发展奠定了基础。为了发展我国独立的网络空间，IPv6 也正在加紧其部署的步伐和研发，近些年 IPv6 网络规模不断扩张[3]。

IPv6 网络的诞生之初是为了解决地址不足的问题。因此，IPv6 网络并未从根本上杜绝网络攻击的发生。在 IPv6 网络中，DoS 攻击仍然是一种较为常见的网络入侵攻击[4]。传统的网络入侵检测软件 Snort 可以对 DoS 攻击进行检测。文献[5]中的检测规则可以采用蜜罐技术的日志进行补充。但是，使用大量具体规则对网络流量中的 DoS 攻击进行检测，会降低 Snort 这类入侵检测软件的执行效率。文献[6]为了解决特征选择的问题，提出了特征选择框架，采用了支持向量机（Support Vector Machine，SVM）和粒子群算法（Particle Swarm Optimization，PSO），确定了最佳检测特征，并实现了对 IPv6 网络 DoS 攻击检测的效果。

在新的 IPv6 环境下，网络流量数量处于一个快速增加的状态，DoS 攻击也并没有消失。IPv6 网络流量的数量比传统 IPv4 更大，对 DoS 攻击进行检测的时间和计算量也会有所增加。因此，在 IPv6 网络中检测 DoS 攻击的方法应当注重检测的性能，需要对重要的特征进行了选择，以尽可能降低检测过程中的特征维度。网络流量的特征可以用于攻击检测，但不同特征在实际检测过程中的影响程度也是有高低之分的。本文应当采用机器学习的方法来实现 IPv6 网络的 DoS 攻击检测，采用双重降维的方法，不仅对特征进行了选择，还实现了对离散类型子特征进行筛选和聚合，同时通过信息增益率体现出了不同特征的权重，将其运用在 DoS 攻击检测中，以提高检测系统的适应性与检测效率。

2.2 信息增益率

信息增益率起源于 C4.5 算法，被用于决策树的分裂过程[7]。基于信息增益的 ID3 算法被用于决策树算法中，但这会造成更加容易倾向选择取值较多的属性[8]。因此，在有些情况下，使用信息增益作为特征权重的参考指标是不合适的。在决策树中，一般信息增益率越大的属性越靠近根节点，这类属性在分类过程中造成的影响也越大。因此，本文采用信息增益率作为特征“双重降维”的重要评价指标，并将信息增益率作为改进距离的权重，以提高分类算法的效果。

信息增益率是在原有的信息增益的基础上进行进一步的计算。例如，对于数据集 D 中的属性 A，在原来计算信息增益为 $Gain(A)$ 的基础上，需根据式(1)得到分裂信息，再根据式(2)即可得到属性 A 的信息增益率[9]。

$$SplitInfo_A(D)=-\sum_{i=1}^{n}\frac{|D_i|}{|D|}\times\log_2\left(\frac{|D_i|}{|D|}\right) \tag{1}$$

$$GainRatio(A)=\frac{Gain(A)}{SplitInfo_A(D)} \tag{2}$$

其中，D_i 为数据集 D 基于属性 A 中的元素进行划分的子集。

2.3 KNN 算法

KNN 算法作为机器学习中常用的算法之一，该算法具有轻量分类的特点。SVM 算法和随机森林算法在使用的过程中需要先基于样本对分类器进行训练。在网络入侵检测系统检测 DoS 攻击时，这往往需要较高的时效性。具有快速响应能力的 KNN 算法，无需在使用前对分类器进行训练，从而此算法可以更好地用于 DoS 入侵检测。

传统的 KNN 算法的实现过程如下：首先，计算待测点与所有已知类型的样本点的距离；其次，按照距离的远近进行排序；最后，基于最近的 k 个样本点，统计每一个类别的样本点的个数，选取个数最多的一类作为待测点的分类结果。然而，对于小群体样本点，在分类过程中可能会因为 k 值的变化产生分类决策劣势的问题。同时，传统 KNN 算法存在因个数相同造成分类决策出现偏差的问题[10]。另一方面，采用普通的欧氏距离作为度量两点之间的距离的方法，在一定程度上不能体现出不同特征的重要性的高低。文献[11]采用了层次分析法进行特征权重的赋值，但是层次分析法可能会受到一定的主观性的影响。文献[12]基于信息增益率，给出了一种在 KNN 算法中对欧氏距离特征加权的思想。因此，在此次研究中，本文对 KNN 算法的不足进行改进，使改进后的算法更好地用于 IPv6 网络中的 DoS 入侵检测。

3 入侵检测算法

3.1 双重特征选择与降维

基于 IPv6 网络流量，可以从中提取流量特征，并用于描述网络流量的变化情况。网络流量特征可按照类型分为离散型特征和连续型特征。其中，离散特征可分为数字离散特征和类型离散特征。若仅仅对原始特征进行特征选择，这还是远远不够的，在实际运算过程中，参与计算的特征维度数量可能还会呈现爆炸式的增长。在分类算法执行过程中，由于类型离散特征均需要化为 0-1 特征，这会造成特征的整体维度数量会快速增长，从而在一定程度上影响分类检测的性能。正如之前所说，IPv6 网络中的流量在快速增长，应当尽可能对最终参与计算的实际特征数量进行降低。因此，基于 IPv6 网络流量特征，不仅需要对原有的一级特征进行降维，还要针对类型离散特征中的二级“类型子特征”进行降维，以提高后期分类的执行效率。

本文采用信息增益率作为特征重要性的评价指标。在信

息增益率的基础之上，对特征实现“双重降维”的功能。对于数字离散特征，可以直接计算特征的信息增益率。在连续型特征的信息增益率计算过程中，需要先对连续型特征实现离散化，再计算特征的信息增益率[13]，以下过程为连续型特征的离散化的方法。对于属性 A，需将属性值进行排序，得到 $\{a_{(1)},a_{(2)},\cdots,a_{(n)}\}$。通过计算 $V_i=\frac{a_{(i)}+a_{(i+1)}}{2}$ 可以得到该特征在该数据集上的一个划分，$T_1=\{a|a\leqslant V_i\}$，$T_2=\{a|a>V_i\}$。其中，由于 V_i 有 $n-1$ 种取值，因此需要分别计算在不同划分下的信息增益率，并选取最大的信息增益率作为 V 的局部信息增益率阈值。在序列中找到离阈值最近的数值，并将不超过阈值的数值作为分割点，从而找到离散化的分割点。由于本文只需计算特征的信息增益率，因此只将最大的局部信息增益率的阈值作为待测属性的信息增益率，而不再通过寻找具体分割点来计算信息增益率。

通过分别计算一级连续型特征和离散型特征的信息增益率，分别选取阈值 τ_{continue} 和 $\tau_{\text{uncontinue}}$ 作为特征过滤的条件，从而完成第一次特征降维的目标，即选取出高质量一级特征。在第二次降维过程中，主要针对非数字离散型的类型离散的“子特征”实现降维。基于子特征的信息增益率，可以通过相对权重计算得到各类子特征的阈值。例如，对于离散非数字型特征 W 的子特征过滤阈值 τ_W 的计算式如下：

$$BaseGI=\frac{\sum_{X_i}^{Y}\sum_{x\in X_i,x\neq 0}x}{\sum_{X_i}^{Y}|X_i|} \tag{3}$$

$$\tau_W=\min\{BaseGI,\frac{GI_w}{\frac{1}{n}\sum_{i=1}^{n}GI_i}*BaseGI\} \tag{4}$$

其中，Y 为所有参与二级子特征降维属性集合 X_i 的并集，即 $Y=\cup X_i$。其中，X_i 中的元素为信息增益率不为 0 的元素，$|X_i|$ 表示集合 X_i 的元素个数，GI_i 为一级特征 i 通过计算所得的信息增益率。

在上述基础之上，我们应当对子特征进行筛选和聚类，以实现特征的双重降维功能。在二级子特征降维的过程中，只需对子特征较多的特征进行降维，而拥有较少子特征的一级类型离散属性特征无需进行降维。本文采用备选子特征集的思想实现对子特征降维的功能。例如，对于某个一级离散类型特征，它拥有较多的子特征，我们需要对其子特征进行降维。首先，通过计算这个离散非数字类型的所有子特征的信息增益率和其子特征的过滤阈值，将超过阈值的子特征加入备选子特征集 T_W，其余的特征加入落选子特征集 F_W。其次，在 T_W 特征集中，选取信息增益率较高的部分子特征予以保留，然后将其剩下的子特征移动到 F_W 集合中。一般来说，选择保留 T_W 特征集中信息增益率较高的前 *bound_number* 个子特征，若该集合中的子特征个数不足 *bound_number* 个，则按照该集合中的实际特征进行保留，其中 *bound_number* 为参数。最后，将该特征的 F_W 集合中的所有子特征聚合成 0-1 离散特征的 W_{others}，从而实现了离散非 0-1 特征的子特征降维功能，整体过程如算法 1 所示。

算法 1 双重特征降维算法

输入：网络流量特征数据集 X

输出：双重降维的数据集 Y

```
1. while(hasNextFeature()) do
2. IGR=getInformation_Gain_Ratio(feature_x)
3. end while //计算信息增益率
4. threshold=getThreshold(IGRs) //计算阈值
5. Y=Compared(threshold,Features)
6. getThresholdSub(IGR_Subs)
7. while(hasNextSubFeatureSet()) do
8.    T,F=Compared(threshold_sub_X ,Feature_Sub_X)
9.    Y.add(T)
10.      FC=Clustering(F) //聚合特征
11.      Y.add(FC)
12. end while
```

3.2 GR-AD-KNN 算法

在网络 DoS 入侵攻击中，由于部分类型的攻击发生的频数较少，因此能用于训练或检测的数据也较少。另一方面，为了提高 IPv6 网络流量 DoS 入侵检测的性能，在检测时，应当选取轻量级的分类算法，以减小因训练而带来的时间和空间代价。因此，无需提前进行训练的轻量级 KNN 算法符合这上述需求。

传统的 KNN 算法存在“小群体分类劣势”的问题。当 k 值选取不恰当时，算法的敏感度较高，稳定性较差，从而导致在对小群体类型的数据进行分类时可能存在误判的情况。因此，若想将 KNN 算法用于 IPv6 网络中的 DoS 入侵检测，需要对这一缺点进行优化。

原始的 KNN 算法在对样本进行分类时，所采用的核心评价指标是基于数量的。这虽然拥有较快的检测速度，但带来了“小群体分类劣势”的问题。在原始 KNN 算法归类的过程中，分类算法认为在最近的 k 个样本点中，所有样本点的价值一样，从而在分类决策时大样本数量群体会对小样本数量群体带来负面影响。事实上，在分类过程中，与待测样本距离不同的点应当在决策中具有不同的价值。因此，本文采用偏移增量平均距离的思想，对 KNN 决策部分的算法进行优化。

另一方面，在传统的 KNN 算法中，通常使用欧氏距离作为衡量两个点远近程度的指标。基于 3.1 节的工作，可以计算得到一级特征和二级子特征及其聚合的子特征属性的“信息增益率”。采用文献[12]的思想，利用信息增益率来优化算法的欧氏距离计算方式，以提高重要特征及其子特征在分类决策时的影响力。因此，在距离定义方面，所有的样本点与待测点的加权距离的定义如下[12]：

$$D(x,Y)=\sqrt{\sum w_i(x_i-Y_i)^2} \tag{5}$$

其中，w_i 为属性 i 的基于信息增益率而得到的权值。

同时，本文在归类决策方面对 KNN 算法进行了优化。改进后的算法不再基于个数来对待测点的样本进行归类，而是在平均距离的基础之上加入了偏移增量的概念，从而得到了改进后的 GR-AD-KNN 算法。

GR-AD-KNN 算法采用了偏移增量的思想，从而可以使得不同远近距离的样本点产生不同的影响力。其中，反向影响力的定义如下：

$$OPInfunence(x,y)=D(x,y)+\Delta(x,\hat{y}) \tag{6}$$

我们用“反向”影响力 $OPInfunence(x,y)$ 来评价不同距离的样本点对待测点的影响力。这种影响力 $OPInfunence(x,y)$ 可以看作是一种负向影响力；数值越大，在分类决策过程中所带来的不利影响就越高。其中，$\Delta(x,y)$ 为影响偏移距离。为了使该计算方法能够体现距离所带来的影响力，具体

实现公式如下：

$$OPInfunence(x,Y_i)=\frac{1}{N_{s\,y_i}}\sum_{s\,y_i\in Y}^{N_s}\left[D(x,y_i)+\frac{D^2(x,y_i)}{D_{\max}(x)}\right] \quad (7)$$

其中，$D_{\max}(x)$为样本到待测点第 k 个远的距离，即：

$$D_{\max}(x)=D(x,Y_{(k)})=\sqrt{\sum w_i(x_i-Y_{(k)^i})^2} \quad (8)$$

若存在 $D_{\max}(x)=0$ 的情况，说明前 k 个点与待测点的各特征均一致，则取 $\Delta(x,y_i)=0$，即 $\frac{D^2(x,y_i)}{D_{\max}(x)}=0$。

最后，基于前 k 个近的样本点，分别计算待测点与每一类样本点的平均“反向”影响力，选取平均“反向”影响力最小的类别为待测点的分类结果。在此算法中，随着与样本点距离的增加，“反向”影响力的增长速度率也增加，造成负面影响程度加速增大。

证明：(1)在目标 x_0 的分类过程中，任取点 A，并令 $D(A,x_0)=d$，对于 $\forall\Delta d,\varepsilon>0$，$\exists$ 点 A'，点 B，点 B'，点 R，S. t.

$$D(A',x_0)=d+\varepsilon \quad (9)$$

$$D(B,x_0)=d+\Delta x \quad (10)$$

$$D(B',x_0)=d+\Delta x+\varepsilon \quad (11)$$

$$D(R,x_0)=D_{\max}(x) \quad (12)$$

那么，考虑函数 $G(x)$的变化规律：

$$G(x)=OPInfunence(x_0,y_{x+\Delta x})=OPInfunence(x_0,y_x)$$

令：

$$F_1=OPInfunence(x_0,B)-OPInfunence(x_0,A)$$

$$F_2=OPInfunence(x_0,B')-OPInfunence(x_0,A')$$

则有：

$$F_1=\Delta x+\frac{[\Delta^2 x+2d\Delta x]}{D_{\max}(x)} \quad (13)$$

$$F_2=\Delta x+\frac{[\Delta^2 x+2d\Delta x+2\varepsilon\Delta x]}{D_{\max}(x)} \quad (14)$$

可以计算得到：

$$F_2-F_1=\frac{2\varepsilon\Delta x}{D_{\max}(x)}>0\Rightarrow F_2>F_1 \quad (15)$$

且有：

$$\Delta G(x)\propto\varepsilon \quad (16)$$

证明完毕。

综上所述，在待测点与样本点的距离等距离增长的条件下，反向影响力的增量的变化量不再是等量增长；距离待测点越远的距离间隔所产生的反向影响力的增量越大。因此，距离越远的样本点的反向影响力的增加速度也就越大。这就需要参与分类决策过程的样本点的“行为”要有整体一致性；因为，距离较远的点会因较大的负面影响对该类的分类产生更“严重”的反向影响力。与传统普通平均距离算法相比，上述决策方法加入了每一类样本点集“整体决策”的思想。

4 实验与结果

4.1 实验环境

IPv6 网络中的 DoS 入侵检测所选取的实验数据集源于 KDD CUP 99[14] 数据集的 10%测试集和训练集，并选取了正常类型样本和与 DoS 攻击相关的攻击类型样本。该数据集中的特征可分为 TCP 基本特征、TCP 内容特征、基于时间和基于主机的网络流量特征；因此，本文用该数据集来测试算法的性能。此外，通过分析，数据集中的离散型特征数量为 9，连续型特征数量为 32。由于在网络中可能部分攻击类型的样本数量较少，在实验过程中，本文保留了全部的小数量样本群体；基于所占的近似百分比，对近似比例进行微调，从大样本群中随机抽取不同数量的样本，作为本次实验的数据集。为了验证算法的分类效果，在测试集的组建过程中，删减了原先训练集中不存在的样本类型，因此本次实验的训练集与测试集的组成如表 1 所列。

表 1 训练集与测试集的构成

Table 1 Constituent of training set and test set

攻击类型	训练集的样本数量	测试集的样本数量
Normal	2000	2100
Smurf	5800	5900
Neptune	2200	2000
Back	2203	1098
Pod	264	87
Teardrop	979	12
Land	21	9

实验主要分为两个部分，第一部分是实现特征双重降维和一级特征和二级特征的信息增益率的计算。在计算一级特征的信息增益率来实现降维功能时，本文采用连续型特征和离散型特征各自的平均值作为过滤阈值。在二级非 0-1 离散子特征降维和阈值计算过程中，只对针对子特征数量较多的特征进行降维与聚合操作。在上述数据集中，特征 Service 和 Flag 属于类别离散特征，并且它们的子特征数量均较多，分别为 70 和 11。因此，只考虑了 Service 和 Flag 的子特征的降维过程；而特征 Protocol_type 只拥有 3 个子特征，因此不需要参与第二次降维的过程。

第二部分实验是为了评价 GR-AD-KNN 算法的性能。在分类算法实验对照设置方面，本文将仅对欧氏距离加权优化的 GR-KNN 算法与 GR-AD-KNN 算法进行实验对比。分别进行 10 次实验，并记录 10 次实验的平均 F1-Score，用于评判算法的优化情况。在 k 值的选取方面，则是设置了 6 组横向对照实验。为了检测样本点的选取对实验分类结果的整体提升效果，通过对比传统基于平均距离决策的 TAD-KNN (Traditional Average Distance-KNN)算法与 GR-AD-KNN 算法的分类能力，来统计检测能力较弱的攻击类型的检测结果，以评价算法的分类性能的提升效果与稳定性。在评价指标方面，我们通过 F1-Score 指标来综合衡量算法的检测性能。其中，F1-Score 指标的计算方式如下：

$$F1\text{-}Score=2\cdot\frac{precision\cdot recall}{precision+recall} \quad (17)$$

其中，*precision* 为精准率，*recall* 为召回率。

4.2 实验结果

基于 4.1 节所设计的实验，首先对实验数据集进行预处理，从而将具有 41 个维度的流量特征转化为 122 维度的特征。通过实现一级降维，特征的维度减少到了 106 个。然后，我们再对 Service 和 Flag 的子特征降维。我们将二次降维过程中的参数 bound_number 的大小设定为 5，最终可将最终参与分类计算的特征维度确定为 36 个。通过计算得到的信息增益率可以作为不同特征的权重。

基于算法 1，Service 和 Flag 特征的二次降维会产生各自类别的 others 聚合子特征，即子特征 *service_others* 和子特征 *flag_ohters*。这两个特征在欧氏距离中的计算权重是通过计算其所在类未被选中且信息增益率非 0 的子特征的信息增益率的平均值所得到的。

本文对 GR-KNN 算法与 GR-AD-KNN 算法进行了对照实验，并设置了 k 分别为 5，10，15，20，25，30 的 6 组横向对照实验，并纪录了实验结果。同时，在算法性能对比测试实验开始前，分别对两个数据集中的数据进行归一化的处理。具体实验比对结果如图 1 所示。

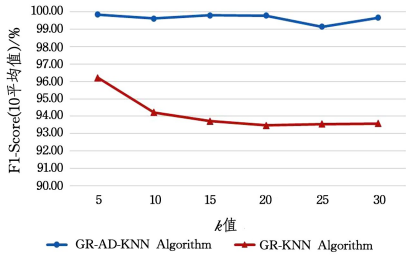

图 1 GR-KNN 算法与 GR-AD-KNN 算法的检测结果

Fig. 1 Detection result of GR-KNN algorithm and GR-AD-KNN algorithm

由图 1 可知，GR-AD-KNN 算法的实验分类结果较好。同时，GR-AD-KNN 算法对 k 值的敏感程度较低，从而降低了算法对 k 值的选择严格度，降低了因模型参数调整时带来的灵敏度过大的负面影响。

另一方面，本文对传统基于平均欧氏距离的 TAD-KNN 算法与 GR-AD-KNN 算法的检测性能进行了对比。当 k 为 5 时，我们分别独立进行了 3 轮实验，每轮实验包括 10 次检测实验。最终，统计每一轮小群体攻击类型为 Teardrop 的 F1-Score 的 10 次检测结果的平均值，如图 2 所示。

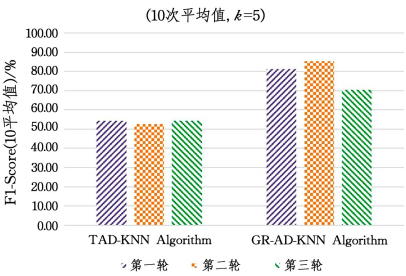

图 2 Teardrop 攻击的 10 次平均 F1-Score 检测结果

Fig. 2 Detection result of 10 times average F1-Score in Teardrop attack

通过实验可以发现，在比较两种算法检测 Teardrop 攻击类型的性能时，GR-AD-KNN 算法对于 Teardrop 攻击类型拥有更好的检测效果，从而可以说明优化后的算法可以提升原先检测能力较弱的攻击类型的检测结果。因此，经过优化后的 KNN 算法用于 IPv6 网络中的 DoS 入侵检测有着更优的分类结果和检测性能。

结束语 本文以 IPv6 网络 DoS 入侵攻击技术为研究背景，以信息增益率为网络特征质量的评价指标，采用双重降维的方法实现了对特征降维的效果，提高了后期分类算法的分类执行效率。在分类算法方面，本文将 KNN 算法的改进算法 GR-AD-KNN 算法运用在了 DoS 攻击检测领域。基于信息增益率所提供的权重，实现了使不同特征产生不同影响程度的效果。基于偏移增量平均距离的思想，改进了待测点的决策算法，实现了不同远近距离的样本点对决策影响力不同的效果，并提高了算法的稳定性，同时缓解了“小群体分类劣势”的问题。因此，上述研究对实现 IPv6 网络中的 DoS 入侵检测有较强的理论意义。

参 考 文 献

[1] SUN S Y. IPv6: Opportunities for the Development of Next Generation Internet in China [J]. The Internet Economy, 2018 (8): 20-25.

[2] PRAPTODIYONO S, MURUGESAN R K, HASBULLAH I H, et al. Security mechanism for IPv6 stateless address autoconfiguration[C]// International Conference on Automation, Cognitive Science, Optics, MICRO Electro-Mechanical System, and Information Technology. IEEE, 2016: 31-36.

[3] WANG J S, LI J Y, ZHANG H W, et al. Design of Large-scale Network Anomaly Traffic Detection System Based on IPv6[J]. Computer Engineering, 2018, 44(10): 14-21.

[4] ARD J B. Internet Protocol version Six (IPv6) at UC Davis: Traffic Analysis with a Security Perspective[J]. Dissertations & Theses - Gradworks, 2012: 20.

[5] SAGALA A. Automatic SNORT IDS rule generation based on honeypot log[C]// International Conference on Information Technology and Electrical Engineering. IEEE, 2016: 576-580.

[6] ZULKIFLEE M. A Framework of Features Selection for IPv6 Network Attacks Detection[J]. WSEAS Transactions on Communications, 2015, 14(46): 399-408.

[7] LI D, LI Y, YUAN C, et al. The application of decision tree C4.5 algorithm to soil quality grade forecasting model[C]// IEEE International Conference on Computer Communication and the Internet. IEEE, 2016: 552-555.

[8] LI B, CHEN A B, ZHOU T, et al. Grade evaluation based on improved C4.5 algorithm in forest fire danger[J]. Hunan Forestry Science & Technology, 2018, 45(1): 36-40.

[9] DU J L, YAN W L. Multiple classifiers of C4.5 decision tree based on distance weight[J]. Computer Engineering and Design, 2018, 39(1): 96-102.

[10] XIAO H H, DUAN Y M. Improved the KNN Algorithm Based on Related to the Distance of Attribute Value[J]. Computer Science, 2013, 40(S2): 157-159, 187.

[11] DAI P W, PAN B, WANG Y M, et al. An Improved KNN Algorithm Based on Analytic Hierarchy Process[J]. Journal of Liaoning Shihua University, 2018, 38(4): 87-92.

[12] PENG Y. Precision marketing system of terminal based on Data mining[D]. Nanjing: Nanjing University, 2014.

[13] ZHANG Y Q. Improvement of Algorithm for finding Segmentation Points of Continuous Value of Decision Tree[J]. Computer CD Software and Applications, 2013, 16(23): 116-117.

[14] KDD Cup 1999 Data [EB/OL]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

ZHAO Zhi-qiang, born in 1994, postgraduate. His main research interests include network security and machine learning.

YI Xiu-shuang, born in 1969, professor, is a member of China Computer Federation. His main research interests include next generation internet, network security and big data analysis.