文章编号:1671-4598(2018)10-0165-05 DOI:10.16526/j.cnki.11—4762/tp.2018.10.036 中图分类号:TN911 文献标识码:A

基于S—Kohonen的DoS攻击检测算法研究

卢 鹏[1], 唐 超[2]
(1. 广州城建职业学院 信息工程学院, 广州 510925;
2. 广州工商学院 计算机科学与工程系, 广州 510800)

摘要: 针对现在DoS攻击检测算法过程中检测效率较低且检测时间比较长的问题, 提出了基于S—Kohonen的DoS攻击检测算法; 使用此算法实现并量化网络流量数据包的分割, 并有效提取累积量特征, 在DoS攻击检测过程中使用累积量; 对现代入侵检测数据集进行全面的分析, 此算法能够实现DoS攻击的全面检测; 与传统以网络流量熵值为基础的异常检测算法相比, 此算法可有效提高检测的精准度, 缩短检测时间。

关键词: S—Kohonen; DoS; 攻击检测

Research on DoS Attack Detection Algorithm Based on S—Kohonen

Lu Peng[1], Tang Chao[2]
(1. Guangzhou City Construction College, Guangzhou 510925, China; 2. Department of Computer Science and Engineering, Guangzhou College of Technology and Business, Guangzhou 510800, China)

Abstract: Aiming at the problem of low detection efficiency and longer detection time in the current DoS attack detection algorithm, it proposed a DoS attack detection algorithm based on S—Kohonen. With this algorithm could realize and quantify the segmentation of network traffic data packets, and extracted the cumulant feature effectively, so as to use cumulants in the process of DoS attack detection. Through the comprehensive analysis of the modern intrusion detection data sets, it would implement full detection of DoS attacks. Compared with the traditional anomaly detection algorithm based on the entropy of network traffic, this algorithm could effectively improve the detection accuracy and shorten the detection time.

Keywords: S—Kohonen; DoS; attack detection

0 引言

拒绝服务(DoS)攻击属于现代最为多见的网络攻击行为, 此攻击的主要目的是打击计算机及网络的正常服务能力, 包括: 1) DoS攻击, 利用对攻击目标发送攻击数据包, 从而有效降低网络及主机的资源, 此种攻击也可以称之为数据包洪泛攻击; 2) 分布式拒绝服务(DDoS), 属于DoS攻击的一种延伸, 其通过因特网分布式连接, 利用控制分布在网络计算机中产生大量数据包洪泛, 实现网络和计算机的攻击[1]。目前入侵检测系统一般使用亡羊补牢工作模式, 也就是在网络和计算机受到多次攻击之后, 通过安全专业人员花费较多时间对攻击数据包进行全面的分析, 并且对攻击特点进行有效总结[2]。但是此种工作模式具有两种缺点: 1) 通过工作人员手工实现攻击特点的总结, 要求消费大量物力及人力; 2) 无法在发生攻击之后及时获取攻击特点实现防御, 不能够降低攻击导致的危害[3]。基于此, 本文研究了以S—Kohonen为基础的DoS攻击检测算法及技术。

1 S—Kohonen网络模型分析

S—Kohonen网络模型指的是没有监督的学习网络, 属于自组织竞争型的神经网络, 能够实现环境特点的自动识别, 还能够实现自动聚类。S—Kohonen神经网络使用自组织特点的映射, 并且对权值进行调整, 在输入模式中通过神经元和网络相互连接, 之后神经元就能够实现攻击的检测[4]。图1为S—Kohonen网络的二维网格结构。

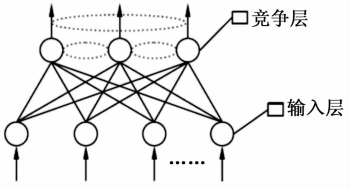

图1 S—Kohonen网络的二维网格结构

S—Kohonen网络在工作过程中的主要原理为: 在样本到网络中输入的时候, 竞争层神经元的计算和输入样本的欧式距离最小的就是获胜神经元, 对相邻神经元和获胜神经元权值进行有效的调整, 使其和周围的权值与输入的样本相互接近。通过反复的训练, 要求同类的神经元能够和权系数相互接近。在学习过程中, 权值的学习速率及神经元的领域都在不断的降低, 以此集中相同的神经元[5]。

S—Kohonen网络有两层, 包括竞争层及输入层, 二维网格节点属于神经元, 输入层具有多个输入节点, 使用j进行表示。S—Kohonen神经网络通过神经元实现拓扑结构的创建, 神经元网络位置之间的联系和其关系具有密切的

收稿日期:2018-02-24; 修回日期:2018-03-23。

作者简介:卢 鹏(1975-),男,湖南桃江人,硕士研究生,高级工程师,主要从事云计算,软件工程,软件测试方向的研究。

联系。

S－Kohonen 算法步骤：

S－Kohonen 神经网络的输出层每个节点都代表一类数据，因此，有多少节点就代表有多少数据类别。输出层与竞争层节点的网络权值互相联系，当在 S－Kohonen 网络中输入训练数据时，输入层、竞争层之间的权值、竞争层、输出层之间的权值需要根据实际应用进行调整。

1）数据归一化。数据的归一化采用函数 mapminmax 处理完成。在进行数据的归一化之前，新建一个 excel 导入需要训练的数据，利用 excel 相关功能对训练数据进行分类，并添加相应的标签，共将训练数据分成 5 类，然后随机将训练数据和测试数据实现排序，排序之后利用公式（1）实现训练数据和测试数据的归一化处理。

$$y = \frac{(y_{\max} - y_{\min}) * (x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min} \tag{1}$$

公式（1）中，$y_{\min}$和 $y_{\max}$均为参数，没有定值，可以自行设置，通常默认为－1，1；

2）初始化。根据输入的数据的具体情况设置输入层、竞争层和输出层节点个数，设置节点个数之后，对竞争层的节点进行排序及设置相关的参数。输入向量 $X(k) = [x_1(n), x_2(n), \cdots, x_N(n)]^T$ 之后作为训练数据。权值向量为 $W_{ij}(k) = [w_{i1}(n), w_{i2}(n), \cdots, w_{iN}(n)]^T, i = 1,2,\cdots,M, j = 1,2,\cdots,N$。令 L 为迭代总次数，设置初始学习速率 α（0），选择邻域半径 N_c（0）。

3）计算优胜节点。d_j 代表输入向量 X（k）$= [x_1(n), x_2(n), \cdots, x_N(n)]^T$ 与竞争层神经元 j 之间的距离。

$$d_j = \left| \sum_{i=1}^{m} (x_i - W_{ij})^2 \right|, j = 1,2,\cdots,n \tag{2}$$

用最小距离输入向量 X（k）的竞争层神经元作为最佳输出神经元。根据公式（2）得到样本的最优节点，也就是和输入样本距离最短的竞争层节点。

4）权值调整。调整包含在其领域 Nc（t）内的节点权系数和节点 c，即：

$$N_c(t) = (t \mid find(norm(pos_t, pos_c) < r))$$
$$t = 1,2,\cdots,n \tag{3}$$
$$W_{ij} = W_{ij} + \eta(X_i - W_{ij}) \tag{4}$$

公式（3）中，pos_t 代表神经元 t 的位置，pos_c 代表神经元 c 的位置；$norm$ 计算不同神经元间的欧式距离；r 为领域半径；η 为学习速率，η 随着进化次数的不断增多而呈现出线性下降的趋势。根据公式（4）对需要的优胜节点权值进行调整，改变学习速率和领域半径的值，使它们随着进程逐渐减小，从而不断拉近输入数据与节点的距离，最终完成神经网络的聚类功能。

5）判断算法是否结束，如果没结束则返回 3）。

6）分类检测。随机对测试数据进行排序，再利用 mapminmax 函数实现归一化处理，数据处理好之后传输至训练好的 S－Kohonen 神经网络实现分类检测。

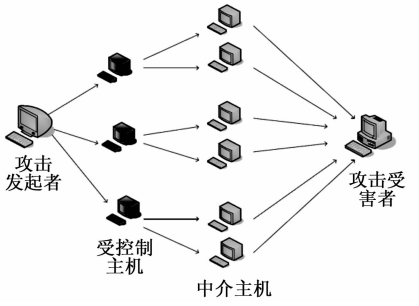

图 2 DOS 的攻击过程结构

2 DoS 的攻击检测算法

DoS 攻击方法在不断的发生变化，通过图 2 可以看出来，攻击发动人员对大量网络防御脆弱主机进行了控制，此主机没有防火墙的安装，还存在一定的软件安全漏洞，这致使其容易被攻击发动人员所控制，控制网络通过伪造源地址等多种方法还能够对受害的主机实现攻击流量的发送，另外，还可借助反射技术对其他的主机朝着目标主机进行攻击[6]。

在 DoS 攻击过程中，一般使用三种手段：第一种，使用虚假源 IP 地址；第二种，利用对僵尸网络进行控制，从而对目标主机进行请求的发送；第三种，实现反射攻击。完整 DoS 攻击主要由多种或者一种项目构成。

图 3 为分布的网络流量目的 IP 地址，通过图 3 可以看出来，在网络中，大部分的目的 IP 少次出现，并且分布并不均匀，以此就出现了网络流量目的 IP 地址重尾特点，也就是大量属性值只是出现了很少的次数，其中的源端口、源地址及目的端口都具有相同的重尾分布规律。网络流量分布特点能够对网络异常及工具进行有效的检测，考虑到异常会对网络流量端口及 IP 地址分布特点进行有效的检测，所以还要使用熵描述[7]，图 4 为目的 IP 地址熵值的曲线。

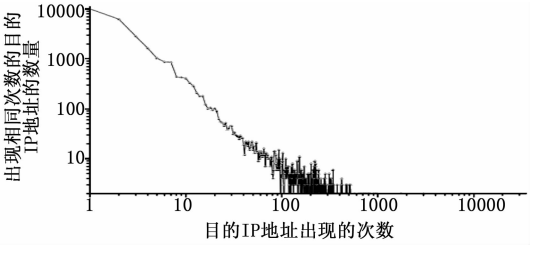

图 3 分布的网络流量目的 IP 地址

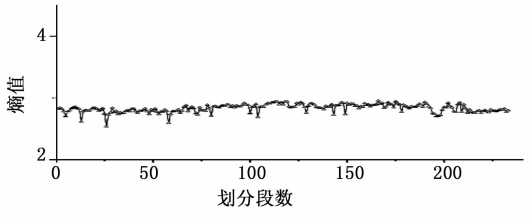

图 4 目的 IP 地址熵值的曲线

表 1 为不同类型异常对于属性分布的特点，通过表 1 可以看出，不同类型攻击都会对目的 IP 地址分布特点产生一定的影响，由此，目的 IP 地址指的是实现网络流量攻击检测的良好属性，为了能够对此影响直观的表现出来，就选择网络入侵检测攻击较多的数据集实现不通过类型攻击对于网络流量的影响[8]，详见图 5。

表 1　不同类型异常对于属性分布的特点

异常	定义	影响
aFlows	短时间产生数据流	目的地址源地址
DoS	拒绝攻击	目的地址源地址
Flash Crowd	对单一地址访问突发流量	目的端口目的地址
Port	扫描大量端口	端口地址
Network	扫描大量主机	地址端口
Outage	通过设备故障导致中断流量	目的源地址
Point	从单源到目的流量	源地址地址
Worms	扫描主机攻击	端口地址

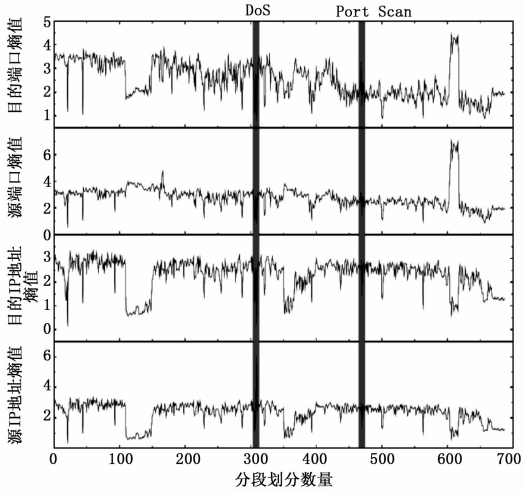

图 5　属性不同熵值的变化曲线

通过图 5 可以看出，在受到 DoS 攻击的时候，目的端口属性就会具有集中性，从而降低目的 IP 地址熵值，而且还会出现大量的虚假源地址，以此分散源端口的属性，提高 IP 地址的熵值。

3　基于 S－Kohonen 的 DoS 攻击检测

3.1　检测的原理

DoS 攻击检测是一种异常检测算法，其主要就是对正常数据进行分析，得到正常数据模型，对需要检测的数据进行判断，从而对异常数据进行确定。因为 DoS 攻击检测使用自适应的检测模型[9]，图 6 为自适应检测模型的工作原理。

在图 6 中，数据收集模块是根据数据收集策略实现网络数据的收集，之后对特征提取模块进行提交。特征提取模块使检测数据能够利用特定算法实现流量特点的转换，流量特点属于检测数据高层的抽象化。攻击检测模块以流量模型的判断是否正常为基础，以此出现检测的结果。自

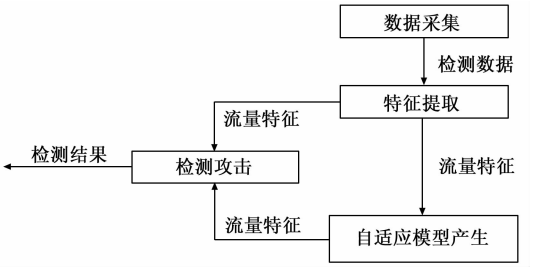

图 6　自适应检测模型的工作原理

适应模型通过被保护网络环境中实现数据的收集，从而产生一开始检测模型，并且在积累大量数据之后，利用自学习能力模型的产生算法产生全新的检测模型[10]。

3.2　检测的过程

DoS 检测的过程主要包括产生检测模型及检测攻击，在产生检测模型过程中，网络中原始数据就会被收集，并且得到相应的处理，而且还产生通过特征创建的聚类检测模型，在检测攻击过程中，就会产生流量特征，之后通过检测模型实现检测。图 7 为检测过程中的检测模型，其属于攻击检测基础。

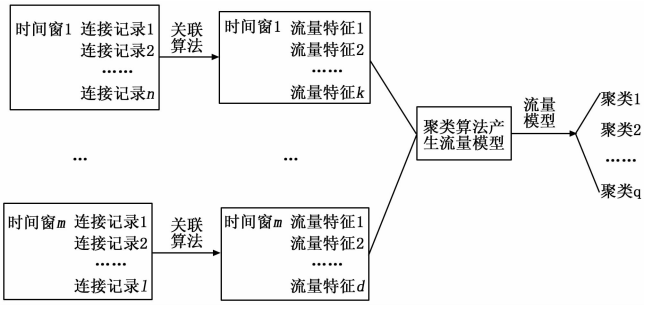

图 7　检测过程中的检测模型

DoS 检测模型主要是将时间窗作为单位处理，实现记录的连接，也就是让网络数据以发生时间为基础，按照时间窗的形式进行相应的划分，使某个时间窗中的原始网络数据实现连接记录的恢复。DoS 在所有时间窗中都实现检测，以此达到实时检测的目的。之后 DoS 通过关联算法使连接记录能够转变成为流量特点，DoS 攻击属于群体网络的行为，从单一 TCP 连接记录来看较为正常，但是在短时间内会出现大量相同 TCP 连接，通过关联算法能够得到此种群体网络行为[11]。

在产生流量特点之后，就要使用聚类算法对正常流量特点聚类进行计算，之后以距离为基础实现异常判断。聚类算法将流量特点作为向量，服务类型等一系列属性属于向量分量。聚类处理结果就是将大量数据组合成为多个数据集合，其中的攻击数据就是小数据集合，详见图 8：

聚类算法能够产生模型，其算法为：

$$d(i,j)=\sqrt{w_1\mid x_{i1}-x_{j1}\mid^2+w_2\mid x_{i2}-x_{j2}\mid^2+\ldots+w_p\mid x_{ip}-x_{jp}\mid^2}$$

其中的每个子集都表示一个聚类，每个聚类中数据距

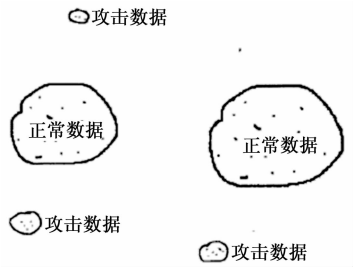

图 8 聚类的模型示意图

离比较近，不同聚类数据间距比较远，不同聚类通过中心值表示，利用聚类数据计算平均值得到中心值[12]。

4 实验结果与分析

为了验证本文方法的有效性，需进行仿真实验，本文实验测试在 Windows 7 操作系统上，处理器为 AMD A10－5750M APU with Radeon（tm）HD Graphics 2.50 GHz，内存为 4.00 GB。实验在 Matlab R2010 b 环境创建网络模型，对 DoS、U2R、Probe、Data、R2L 五种数据进行分析，它们属于现代攻击检测过程中使用最为全面的数据集。

为了分析本文提出的算法的有效性检测 Smurf 攻击。当 Smurf 攻击出现时大量主机会同时向受害主机发送 ICMP 报文，由于该段时间内源地址数目会显显著增多，因此，会使系统的原有秩序会被打乱，使某一时段目的 IP 地址数目相对集中，产生新的大量的源 IP 地址或者原有的大量目的 IP 地址消失，影响受害主机的 IP 地址出现概率，最终改变网络流量的分布结构。

目前攻击过程检测方法中熵值的使用最为广泛，其属于对网络特点的有效描述，也是不确定特点的度量。熵值越大，随机性分布会越明显。熵值越小，便更加表现为集中式分布。在实验过程中使用三折较差验证，使样本集数据分成三组，每个子集数据中具有一次验证集，其他的子集数据为训练集[13]。

网络 Smurf 攻击的采集数据样本来自于 KDP 网络病毒数据库，采集相关的攻击数据构成测试集，在采样过程中测试集的采样率为 $fs=10$ kHz，网络攻击数据的特征分解带宽 $B=1000$ Hz。自相关匹配滤波器的参数为：初始步长 $\mu_0=0.001$，$\theta_2=0.45\pi$，$\theta_1=-0.3\pi$，即 500 Hz。根据上述仿真环境和参数设定，得到的四组实验结果具体如下：

第一组实验中的结果详见图 9，通过图 9 可以看出，在收集窗口一样的背景下，不同比例滑动窗口对于 ABA 算法检测性能不存在影响，并且具有一定的算法稳定性。该算法的检测率不高，处于 90%～90.5%之间。

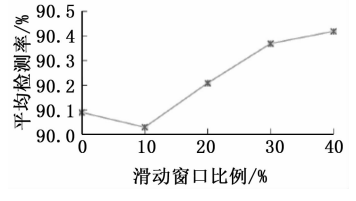

图 9 不同滑动窗口的平均检测数

第二组实验的结果详见图 10，通过图 10 可以看出，不同采样窗口对于算法都具有不同的影响，采样窗口越大，平均检测率越低，因为攻击突发网络信号流量在短时间中具有明显的抖动，影响了积累量特征值，对检测率有所降低[14]。

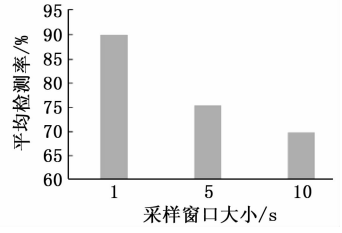

图 10 不同收集窗口中算法的平均检测数

第三组实验的结构详见图 11，通过图 11 可以看出，在 1 s 时间窗口中的 sport 检测率是最高的，在 5 s 和 10 s 时间窗口中的 dip 检测率最高。

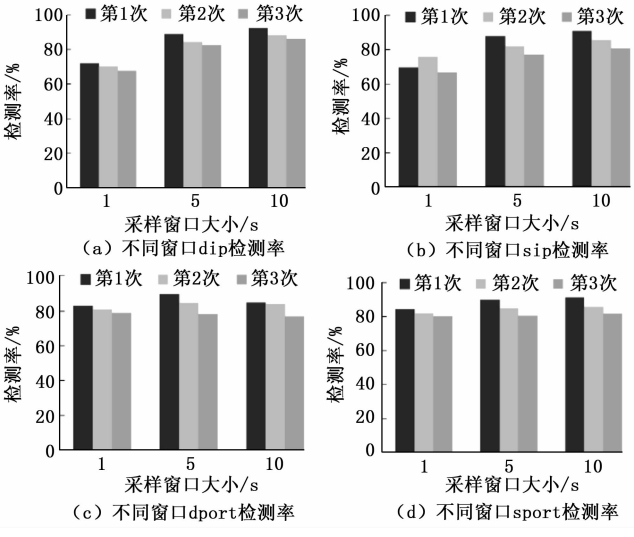

图 11 第三次的实验结果

通过以上分析可以看出，本文所研究的算法具有较高的精准度，正确率和运行时间上都较优，能够为网络管理工作人员提供可靠的响应时间，从而有效阻止网络攻击，具有良好的分类检测性能。

5 结束语

本文对 S－Kohonen 神经网络流程及缺点进行了全面的研究，并且实现了算法的优化，之后创建模型，能够有效提高攻击检测算法的检测效率。最后，本文使用试验验证的方法进行分析，通过结果表示，基于 S－Kohonen 网络的 DoS 攻击检测算法能够有效提高分类正确率，并且缩短训练时间，提高模型精度。

参考文献：

[1] 陈 东. 无线传感器网络中 Path－based DoS 攻击检测算法和防御机制研究 [D]. 北京：北京交通大学，2015.

[2] 李昆仑，董 宁，关立伟，等. 一种改进 Kohonen 网络的 DoS

攻击检测算法［J］. 小型微型计算机系统，2017，38（3）：450－454.
[3] 江　超. 无线传感器网络中基于免疫原理的 DoS 攻击检测算法［J］. 传感器与微系统，2013，32（1）：141－144.
[4] 王秀娟，相从斌. 基于累积量的 DoS 攻击检测算法［J］. 北京工业大学学报，2017，43（9）：1328－1334.
[5] 张晓瑜，吴志军，岳　猛，等. 基于网络流量奇异性特征的 LDoS 攻击检测方法［J］. 计算机工程与设计，2016，37（1）：50－54.
[6] 赵　康，武　斌，范双娇，等. 基于 SVM 的无线局域网 DoS 攻击检测［A］. 全国青年通信学术年会［C］. 2014.
[7] 许学添. 基于幅频响应带宽检测的网络 DOS 攻击识别算法［J］. 信息通信，2016，21（9）：223－225.
[8] 谢柏林，蒋盛益，张倩生，等. 基于 HMM 的应用层 DoS 攻击检测方法［J］. 计算机应用研究，2013，30（11）：3393－3395.
[9] 楼恒越，窦　军. 一种针对基于 OpenFlow 的 SDN 网络中控制层面的 DoS 攻击研究［J］. 计算机科学，2015，15（b11）：341－344.
[10] 刘衍珩，付　枫，朱建启，等. 基于活跃熵的 DoS 攻击检测模型［J］. 吉林大学学报（工），2011，41（4）：1059－1064.
[11] 罗　捷，武　斌，沈焱萍. 基于优化 PSO－BP 算法的无线局域网 DoS 攻击检测［A］. 中国通信学会学术年会［C］. 2014.
[12] 陈世文. 基于谱分析与统计机器学习的 DDoS 攻击检测技术研究［D］. 郑州：解放军信息工程大学，2013.
[13] 朱　勇，罗军舟，李　伟，等. 一种基于概率滑动窗口的应用层 DoS 攻击防御模型［J］. 解放军理工大学自然科学版，2012，13（1）：34－40.
[14] 刘　鹏，孔　宁，田　野，等. PCacheDS：一种基于主动缓存算法的发现服务［J］. 计算机应用研究，2016，33（4）：1172－1178.

（上接第 159 页）

5　总结与展望

随着通信环境日渐恶劣、通信需求日益提升，研究新型的抗干扰技术及策略成为重要议题。高速抗干扰波形是一个重要的研究方向，新兴技术 NC－OFDM 与 TDCS 都可以通过频谱感知模块达到剔除不可用子载波的目的，其中 NC－OFDM 可以实现较小较弱窄带干扰情况下的高速数据传输，TDCS 可以实现较大窄带干扰情况下速率高于扩频的可靠传输，而传统扩频技术则可以满足无频谱空穴情况下的数据传输。通过切换三种模式的不同波形，能够在某些干扰情况下显著提高频谱利用率，改善通信质量。基于此目的，本系统针对电磁环境进行实时波形切换，以此弥补单一模式、单一波形对环境适应性不足的问题，在提升频谱利用率的同时，也提升了系统的鲁棒性，实现频谱资源、系统资源利用的最大化。

由于各个频点处的能量具有随机性、分散性与复杂性的特点，而本次研究中的干扰分类算法目前较为理想，因此需要对干扰分类算法进行优化，为了改善这种情况，可以使用一维卷积神经网络[15]来对干扰进行识别，再根据分类结果对波形决策算法进行优化。由于梳状干扰的形态复杂，针对梳状干扰的波形决策算法也难以在复杂情况下分析最优波形，因此对梳状干扰的波形决策算法的设计有待改善，比如对梳状干扰进行进一步分类。

随着计算能力与日俱增引起的人工智能崛起，必将成为通信领域的一大助力，无论是针对不同干扰选择最佳的通信波形，或是针对不同干扰进行相应的抵消，还是更高阶的实时波形创造，都是基于人工智能与模式识别的大时代赠与我们解决通信问题的一种新思路。通过人工智能算法与先进抗干扰波形相结合，从而实现通信领域中抗干扰的宽带化与智能化，是人工智能与通信技术相互融合进化一个必然方向。

参考文献：

[1] 郭彩丽，张天魁，曾志民，等. 认知无线电关键技术及应用的研究现状［J］. 电信科学，2006，8：50－55.
[2] 夏永平，陈自力，周子栋. 基于 USRP 的变换域通信系统抗干扰平台实现［J］. 军械工程学院学报，2017，29（2）：75－78.
[3] 李少谦，程郁凡，董彬虹，等. 智能抗干扰通信技术研究［J］. 无线电通信技术，2012，38（1）：1－4.
[4] 王　凯，徐展琦，肖永伟，等. 无线认知开发平台综述［J］. 无线电通信技术，2016，42（2）：9－11.
[5] 陈旭东，陈章进，李翰超，等. 基于 FPGA 的频谱分析系统研究与实现［J］. 电子测量技术，2016，39（11）：113－117.
[6] 刘小玲. NC－OFDM 系统的智能抗干扰决策技术研究［D］. 电子科技大学，2016.
[7] 王柔溪，变换域通信系统的关键技术研究与仿真分析［D］. 北京：北京理工大学，2016.
[8] 柯英豪，基于变换域通信系统的改进与实现［D］. 电子科技大学，2016.
[9] 谢铁城，达新宇，褚振勇，等. CCSK 调制的变换域通信系统基函数序列估算算法［J］. 系统仿真学报，2014，26（8）：1713－1717.
[10] 谢铁城，达新宇，褚振勇，等. 采用时频分析的变换域通信系统基函数设计［J］，西安交通大学学报，2012，46（6）：42－47.
[11] 范　伟，翟传润，战兴群. 基于 MATLAB 的扩频通信系统仿真研究［J］. 微计算机信息，2006，22（7）242－244.
[12] 赵建功，刘香玲，朱行信，等. MSK 扩频调制的带限滤波技术分析［J］. 无线电工程，2016，46（1）：65－68.
[13] 王　鑫. 一种基于径向基函数的模型参考自适应控制的研究［J］. 科技创新与应用，2017，26：12－13.
[14] 郑　爽，周冬梅. OFDM 系统 16QAM 和 QPSK 调制仿真分析［J］. 电脑知识与技术，2013，9（15）：3606－3609.
[15] 周飞燕，金林鹏，董　军. 卷积神经网络研究综述［J］. 计算机学报，2017，40（6）：1229－1250.