计算机科学与探索
Journal of Frontiers of Computer Science and Technology
ISSN 1673-9418,CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：多通道自编码器深度学习的入侵检测方法
作者：杨杰，唐亚纯，谭道军，刘小兵
网络首发日期：2020-10-07
引用格式：杨杰，唐亚纯，谭道军，刘小兵. 多通道自编码器深度学习的入侵检测方法. 计算机科学与探索.
https://kns.cnki.net/kcms/detail/11.5602.TP.20201005.1535.002.html

计算机科学与探索
Journal of Frontiers of Computer Science and Technology doi: 10.3778/j.issn.1673-9418.2007023

多通道自编码器深度学习的入侵检测方法

杨　杰，唐亚纯+，谭道军，刘小兵

湖南科技学院电子与信息工程学院，湖南 永州 425199
+ 通信作者 E-mail:tyc770880@sina.com

摘　要：针对现有的入侵检测方法在检测准确率和误报率方面存在的不足，提出了一种多通道自编码器深度学习的入侵检测方法，该方法分为无监督学习和有监督学习两个阶段：首先分别采用正常流量和攻击流量训练两个独立的自编码器，其重构的两个新特征向量与原始样本共同组成多通道特征向量表示，然后利用1维卷积神经网络(convolution neural network，CNN)对多通道特征向量表示进行处理，学习通道之间可能的依赖关系，用于更好地区分正常流量和攻击流量之间的差异。所提方法将无监督的多通道特征学习和有监督的跨通道特征依赖学习有机地结合起来，用于训练灵活有效的入侵检测模型，达到极大地提高模型检测准确率的目的。同时，为了优化 CNN 的超参数并提高网络对通道间依赖关系的辨识效果，利用遗传算法自动寻找 CNN 模型的最优拓扑集合。实验结果表明，所提方法在多个数据集中获得良好的结果，比其他入侵检测算法具有更好的预测准确性。

关键词：入侵检测；自编码器；深度学习；多通道；遗传算法
文献标识码：A　中图分类号：TP393

杨杰，唐亚纯，谭道军，等. 多通道自编码器深度学习的入侵检测方法[J]. 计算机科学与探索
YANG J, TANG Y C, TAN D J, et al. Intrusion Detection Method of Multi-Channel Autoencoder Deep Learning[J]. Journal of Frontiers of Computer Science and Technology

Intrusion Detection Method of Multi-Channel Autoencoder Deep Learning

YANG Jie, TANG Yachun+, TAN Daojun, LIU Xiaobing

School of Electronics and Information Engineering, Hunan University of Science and Engineering, Yongzhou, Hunan 425199, China

Abstract: Aiming at the shortcomings of the existing intrusion detection methods in detection accuracy and false alarm rate, an intrusion detection method of multi-channel autoencoder deep learning is proposed. The method is

*General Project of Hunan Natural Science Foundation No. 2018JJ2147 (湖南省自然科学基金面上项目); Youth Project of Hunan Natural Science Foundation No. 2018JJ3203 (湖南省自然科学基金青年项目); Project of Hunan Science and Technology Department No. 2019ZK4018 (湖南省科技厅计划项目); Hunan University of Science and Engineering Computer Application Special Subject Funding (湖南科技学院计算机应用特色学科).

divided into two stages: unsupervised learning and supervised learning. Firstly, two independent autoencoders are trained by normal traffic and attack traffic respectively, and the two new feature vectors reconstructed and the original samples form a multi-channel eigenvector representation. Then, the 1-D convolution neural network (CNN) is used to process the multi-channel eigenvector representation, and the possible dependence between channels is learned to better distinguish the difference between normal traffic and attack traffic. The proposed method combines unsupervised multi-channel feature learning with supervised cross-channel feature dependence learning to train a flexible and effective intrusion detection model, which greatly improves the accuracy of model detection. At the same time, in order to optimize the parameters of CNN and improve the identification effect of network on channel dependence, genetic algorithm is used to automatically find the optimal topology set of CNN model. Experimental results show that the proposed method achieves good results in multiple data sets and has better prediction accuracy than other intrusion detection algorithms.

Key words: Intrusion detection; Autoencoder; Deep learning; Multi-channel; Genetic algorithm

1 引言

随着互联网、人工智能和大数据技术的飞速发展，网络安全面临着前所未有的复杂威胁。当前人们需要更强大、更有效的网络入侵检测系统(network intrusion detection system，NIDS)来检测日益多样而复杂的网络入侵行为[1]。网络入侵检测系统的目标是通过分析网络流量中是否存在恶意活动来发现计算机网络中任何未经授权的访问[2]，其任务是建立一个能够区分攻击和正常网络流的预测模型。因此，NIDS可以将入侵检测转化为模式识别和分类，利用相关算法对网络中的各种行为进行收集、清理、建模和分类[3]。

经过几十年的发展，目前 NIDS 方法基于检测方式的不同，可以分为基于特征检测的方法和基于异常检测的方法[4]。基于特征检测的方法[5]首先对各种攻击模式进行分析，提取攻击特征，然后将特征加入到特征库用于检测新的攻击，当检测样本与特征库信息匹配时既可判定为攻击行为。这类方法误报率低，但具有一定的滞后性，只能检测特征库内已有的攻击模式，对新型攻击的检测率低。基于异常检测的方法[6]通过建立概率统计模型，对正常信息流量样本的模式进行泛化描述并生成基准模型，当检测样本的模式与基准模型不符时既可认定为异常攻击。该类方法漏报率低，对新型攻击具有一定的检测能力，但是也存在检测精度不够的问题。针对当前入侵检测方面面临的问题，许多研究学者提出了多种技术用于提高检测方法的准确度和稳定性，其中大多数方法是通过降低误报率和检测未知攻击来提高检测精度[7]。随着深度学习的出现和发展，手工定义特征的任务能够被可训练的多层网络所取代，从而在入侵检测任务中能够获得较传统机器学习更高的准确性和更低的误报率，因此各类深度学习方法广泛应用于 NIDS 领域。Lin 等[7]提出了一种基于卷积神经网络(*convolutional Neural Networks，CNN*)方法来进行网络入侵检测的方法，该方法随着训练样本数目的增多而精度随之提高。Almiani 等[8]提出了一种基于多层递归神经网络的自动入侵检测系统用于抵御雾计算面临的网络攻击。张思聪等[9]提出了一种基于深度卷积神经网络(*deep convolution neural network，dCNN*)的入侵检测方法，该方法最大的特点是将一维入侵数据转换为二维"图像数据"用于训练网络，有效提升了入侵检测系统的检测精度。虽然准确率和误报率一直是 NIDS 研究的重点，但是实

时性和检测效率也是一项很重要的指标。在深度学习中，相较于其他神经网络，自编码器能够有效地降低特征维度，并容易与其他模型相结合，在提高检测精度的同时，大大降低模型的训练时间。自编码器作为一种无监督学习的人工神经网络，由一个将输入映射到一个隐藏层的编码器函数和一个解码器函数组成，通过最小化损失函数来产生所学习的重构输入。Alqatf 等[10]提出了一种基于堆叠自编码器的深度学习框架，用于正常样本的特征学习，然后使用支持向量机(*support vector machine, SVM*)分类器提高方法的准确性。Andresini 等[11]采用深度神经网络训练输入数据的特征，然后将自编码器作为异常检测器用于细化分类，从而提高了对不可预见攻击的分类精度。Mirza 等[12]提出了一种基于稀疏自编码器和长短期记忆网络(*long short-term memory，LSTM*)混合的深度学习网络入侵检测方法，该方法利用自编码器进行数据的降维和特征提取，而后使用 LSTM 网络来处理计算机网络数据的顺序性质，从而能够有效地应对不可预见和不可预测的网络攻击。虽然上述方法能够取得一定的效果，但是，这类方法在利用自编码器压缩原始数据同时，不可避免的要丢失掉信息，而且这些研究仅从正常样本中训练单个自编码器，未考虑攻击样本，因此对新型攻击样本的检测率较低。

为了解决传统 NIDS 对新型入侵行为检测精度低、误报率高以及检测效率低的问题，提出了一种基于多通道自编码器的深度学习方法，该方法结合了基于特征检测和异常检测方法的优点，在无监督阶段分别采用正常流量和攻击流量训练两个自编码器，两个自编码器根据输入样本重构出两个新特征向量，然后，将两个重构特征向量与原始样本共同组成多通道的特征向量表示，最后利用 1 维 CNN 网络对多通道特征向量表示进行处理，学习通道之间可能存在的依赖关系。所提算法将无监督的多通道特征学习和有监督的跨通道特征依赖有效地结合起来，用于提高模型的准确性。

所提方法的创新之处在于使用自编码器派生流量数据的多通道表示形式，分别利用正常流量和攻击流量来训练自编码器模型，它们用于为网络流的原始特征向量表示提供这些自动编码器构建的特征向量。而后借助卷积的深度学习架构，来揭示跨通道特征表示中隐藏的依赖关系，而这个特征表示可能有助于入侵检测模型将攻击流从正常流中分离出来，从而保证所提方法能够有效降低误报率以及提高未知攻击的检测精度。

2 深度学习模型

2.1 自编码器

自编码器[12]是一种经过特殊训练的神经网络，其作用是通过将输入数据作为学习目标，对输入数据进行表征学习。自编码器网络由两个函数组成：编码器 f 通过映射函数 $h=f(x)$ 将输入向量 x 映射到隐藏表示 h，参数为 θ_f，解码器 g 是利用映射函数 $\hat{x}=g(h)$ 将得到的隐藏表示 h 映射回输入空间中，获得重构向量 $\hat{x}$，参数为 θ_g。通常，函数 g 和 f 分别对应不同的神经网络。由两个网络组成的自编码器所包含的参数 $\{\theta_f,\theta_g\}$ 通过最小化损失函数 $L(x,g(f(x)))=L(x,\hat{x})$ 来共同学习，其中损失函数的惩罚项可以定义为 $L_{se}(x,\hat{x})=\|x-\hat{x}\|^2$。

2.2 卷积神经网络

卷积神经网络 CNN 是一类包含卷积计算且具有深度结构的前馈神经网络，主要用于处理具有网格状拓扑结构的数据。CNN 可以通过应用相关过滤器成功捕获数据中的空间或时间依赖性，从而产生良

好的内部表示。通常，CNN 通常由两部分组成：一个是特征提取部分，由交替的卷积层和空间池化层组成；另一个是用于预测或分类的部分，由可训练的完全连接层和分类器组成。

一般来说，当局部特征之间的位置关系被确定时，卷积层提取具有局部相关性的特征。卷积利用多个一定的权重(即卷积内核)，对一个区域内的数据进行内积运算，其输出值就是提取的特征之一。构成用于乘法运算的核元素对应于传统神经网络的权值矩阵。所有可能偏移的可训练核与输入数据卷积，在卷积层生成特征映射。卷积层根据线性卷积滤波器和非线性激活函数构建特征映射 f：

$$f_{i,j,k} = \sigma\left(w_k^T x_{ij}\right) \tag{1}$$

式中：(i,j)表示特征映射中的位置，x_{ij} 表示中心为(i,j)的输入数据，k 表示特征图中的通道索引。

3 基于多通道自编码器的深度学习

本文提出的用于解决网络入侵检测问题的多通道深度学习方法，它结合了基于两个自编码器的多通道特征构建的无监督方法和利用跨通道特征相关性的有监督方法。提出模型的总体架构如图 1 所示。

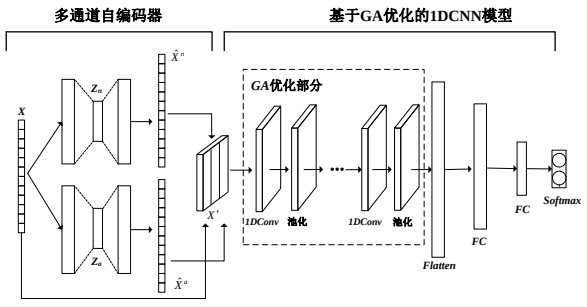


Fig.1 The architecture of the proposed model

图 1 所提模型的架构示意图

3.1 多通道自编码器

本文使用特殊类的自动编码器进行特征学习，将单通道样本转换为多通道样本。假定 $O = \{(x_i, y_i)\}_{i=1}^{N}$ 是一组 N 个训练样本，其中每个 $x_i \in R^D$ 是与 D 个特征上定义的输入样本相对应的行向量，y_i 表示正常样本或攻击样本的对应二进制标签。$X = [x_1, \cdots, x_N]^T \in R^{N \times D}$ 表示 N 个 D 维随机变量 $x \in R^D$ 的数据矩阵。采用 $X^n = X_{|y_i = n}$ 和 $X^a = X_{|y_i = a}$ 分别表示 X 中标记为正常样本和攻击样本的子集。然后，X^n 和 X^a 中的样本可分别用于学习两个独立的自动编码器 $g_n \cdot f_n$(以 zn 表示)和 $g_a \cdot f_a$(以 za 表示)。由于解码器网络顶层产生的激活对应于同一输入空间中的一个重构向量，因此本文思想是将其视为新的学习特征。特别地，每个自动编码器可以被用于从样本 x 构建新的特征向量 $\hat{x} = g(f(x)) \in R^D$，然后这些特征可以被认为通过级联将单通道样本转换为多通道样本。因此，每个样本 $x_i \in R^D$ 都可以由多通道样本代替：

$$x_i' = \left[x_i, \hat{x}_i^n, \hat{x}_i^a\right]^T \in R^{D \times 3} \tag{2}$$

式中：$\hat{x}_i^n = g_n(f_n(x_i))$ 和 $\hat{x}_i^a = g_a(f_a(x_i))$ 分别表示单通道样本 xi 的重构表示。因此，单通道数据矩阵 X 可以扩展为多通道数据矩阵：

$$X' = [x_1', \cdots, x_N']^T \in R^{N \times D \times 3} \tag{3}$$

通过这种方式，将正常样本和攻击样本的自编码器重构的特征进行合成用于丰富样本 x_i 的信息。当样本属于两个不同的分布时，标记为正常的样本 x_i 应该与 $\hat{x}_i^n$ 而非 $\hat{x}_i^a$ 的表示形式更相似，即 $\left\|x_i - \hat{x}_i^n\right\|^2 < \left\|x_i - \hat{x}_i^a\right\|^2$；反之亦然。自编码器步骤中的目标是在受监督阶段利用一个通道对其他每个通道

的影响，以便更好地区分正常和攻击两个类之间的差异。

对于学习三个通道之间的表示有两种方案：一种解决方案是采用(1×1)卷积核的滤波器，该滤波器用于跨通道参数卷积，因此可以通过增加特征数量将学习表示以 $R^{D\times 3}$ (3 通道表示)而不是 R^{3D} 的形式进行连接；另一种解决方案是将学习到的特征连接到空间 R^{3D} 而不是 $R^{D\times 3}$ 中，然后将它们输入到完全连接层中。然而，在完全连接层中，由于训练的输出不受输入特征的顺序的影响，使得输入的拓扑被完全忽略。因此，使用这种方案可能会失去基于信道的排序和学习新的跨信道特性的可能性。

本文定义的自动编码器为三层，这些层中包含 40×10×40 个神经元。为了执行数据正则化并且防止出现过度拟合，在解码层之前放置一个 dropout 层，其定义可以表示为：

$$\begin{cases} r^{l} \sim Bernoulli(p) \\ y^{l'} = r^{l} * y^{l} \\ z_i^{l+1} = w_i^{l+1} * y^{l'} + b_i^{l+1} \\ y_i^{l+1} = f\left(z_i^{l+1}\right) \end{cases} \tag{4}$$

式中：l 表示神经网络层号，r^{l} 表示满足伯努利分布的随机向量，p 表示保留概率，y^{l} 和 $y^{l'}$ 分别表示神经网络第 l 层神经元的输出向量和经过随机阻断后的随机向量。w_i^{l+1} 和 b_i^{l+1} 分别表示第 l 层第个神经元的权重系数和偏置，f 表示激活函数，z_i^{l+1} 和 y_i^{l+1} 分别表示第 l 层第 i 个神经元激活函数的输入值和输出值。自编码器网络训练过程中使用 dropout 是在每次训练过程中按照概率 p 随机将部分神经元的权重置为 0，即将某些神经元丢掉，这样可以缩减参数量，使得局部数据簇差异性更加明显，从而实现避免过拟合。Dropout 层一般选择保留概率 p 的值为 0.5，因为此时 Dropout 层效果最好，生成的网络结构最丰富。

神经网络中常见的激活函数有 Sigmoid、tanh 和 ReLU 等函数。对于每个隐藏层，本文选择整流线性单元(ReLu)作为激活函数，而对于最后一层，则使用线性激活函数 Sigmoid。多通道自动编码器采用均方误差(*mean square error，MSE*)作为损失函数。

3.2 卷积神经网络

由于 CNN 起初被广泛应用于图像处理领域，网络的输入大多数是二维矩阵形式，因此特征图和卷积核等网络内部结构均被设置为二维。随着语言识别领域引入 CNN 后，为了适应语言信号的一维特性，一维 CNN 应运而生。一维 CNN 处理一维输入向量，且卷积中的滤波器仅沿一维滑动。为了更好地利用跨通道的特征组合信息，本文采用一维 CNN 进行特征处理。

对于 1D CNN 模型，为了给过滤器尺寸降维和减少训练过程中的计算量，可以将过滤器的核大小限制为 1，即使用(1×1)的卷积核来减小空间尺寸。同时，(1×1)卷积还具有在通道维度上组合现有信息以获得更多抽象通道信息的效果[14]。本文的想法是利用一维卷积，以增加跨信道信息，即将 3 通道表示 X' 输入到滤波器大小等于 1 的 1D 卷积层。然后使用多于三个的滤波器组，用于增加非线性跨通道依赖的数量。特别地，给定样本 $x_i' = \left[x_i, \hat{x}_i^{n}, \hat{x}_i^{a}\right]^{T} \in R^{D\times 3}$，一维卷积的感受野由 $x'_{i,j,:}$ 表示，并且每个滤波器 k 用于计算特征映射中信号 $f_{i,j,k}$ 的公式定义如下：

$$f_{i,j,k} = \sigma\left(w_k^{T} x'_{i,j,:} + b_k\right) \tag{5}$$

式中：$w_k, b_k \in R^{3}$ 分别表示为权重和偏置系数，σ表示非线性激活函数。$x'_{i,j}$ 中每个特征的通道用相同的共享权重系数进行卷积，并以非线性映射到特征映射。在卷积层中采用 K 个滤波器，可以将 $R^{D\times 3}$ 中

的样本转换为 $R^{D\times K}$ 中的特征映射。然后将一维卷积层的输出展平并作为两个堆叠的全连接层的输入处理。全连接层的作用是将在卷积层和池化层中生成的神经元连接到更高层的所有神经元，并使用最终的分类器模块执行分类操作，输出分类概率。本文使用 Softmax 分类器作为最后的分类模块，在回归中将 x 分类为类别 j 的概率为：

$$s\left(y^{i}=j\mid x^{i};\theta\right)=\frac{e^{\theta_j x^{i}}}{\sum_{k} e^{\theta_k x^{i}}} \tag{6}$$

式中：θ_j 表示第 j 个权重向量，x^i 为第 i 个数据样本。

3.3 GA 优化的 CNN 模型

在各个研究领域，基于 CNN 的深度学习技术已取得了巨大的进步。尽管深度学习可以以数据驱动的方式学习特征和优化权重参数，但是模型架构的选择仍然是一个非常直观的手动过程。输入和输出变量以及学习参数的网络拓扑结构需要人工干预进行调整。然而，由于网络结构的总数随着网络深度的增加呈指数级增长，因此不可能对所有可能的候选结构进行评估来获得最佳的权重参数。为了克服这一繁琐的过程并保持模型的灵活性，本文提出了一种利用遗传算法(*genetic algorithm，GA*)自动寻找 CNN 模型最优拓扑的方法。

本节主要利用 GA 算法关注 CNN 的特征提取部分，通过优化生成最佳拓扑。当内核在原始数据上移动时，CNN 的卷积和池化层会检测模式并提取给定输入的关键特征。因此，为这些层确定适当的权重参数可以提升相当大网络的性能。此类超参数的最佳子集可能会随输入数据的动态变化而变化，因此最好使用系统方法确定这些参数。如果卷积核非常大，网络就不容易考虑输入数据的细节特性，而当内核非常小时，可能会因为学习过多的信息而造成混乱。此外，每个卷积层的核数会影响特征学习的过程，每个内核生成不同的特征映射，以不同的视角充当特征检测器。随着核数的增加，分析输入数据的角度也不同。并且，在增加核数的过程中，还需要找到一个能很好地从输入数据中学习特征并降低计算复杂度的最优值。在大多数使用 CNN 的研究中，网络结构的选择是基于经验表现而不是理论依据。

本文采用遗传算法优化一维 CNN 模型的拓扑结构，以提高入侵检测时的预测性能。由于每个架构因素都会影响网络的性能，因此必须同时调整所有参数才能找到 CNN 的最佳结构。通过 GA 算法对卷积和池化层操作的内核数、内核大小以及池化窗口大小进行优化。将每层组件的大小以二进制字符串编码。如图 2 所示。

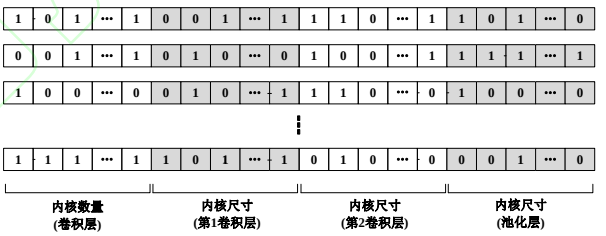


Fig.2 Chromosome structure in CNN optimization model

图 2 CNN 优化模型中的染色体结构

基于 GA 的 CNN 优化过程包括以下步骤：

第一步为初始化种群：遗传算法的搜索过程首先创建一组可能的解。遗传算法最重要的任务是表达染色体所要解决的问题的潜在解，并为每个染色体的性能建立度量标准。由于潜在解的表达影响到所有的遗传操作，每一条染色体都应该恰当地反映目标问题的特征。因此，每个染色体包含每个卷积层内核数、内核大小以及窗口大小对应的潜在值。

第二步为定义适应度函数：通过预定义的适应度函数来测量每个染色体的性能。本文由入侵检测

总体分类准确性作为适应度函数。根据适应度评分，达到较高分类准确度的染色体将比具有较低分类准确度的染色体更频繁地繁殖。

第三步为遗传操作：一旦计算出整个种群的适应度值，就应用选择、交叉和突变的遗传算子来产生下一代的新种群。遗传算子为种群提供了新的信息，通过这些操作，GA 趋向于收敛于最优或接近最优的解决方案。

第四步输出最优解：基于 GA 搜索，获得 CNN 的超参数的最终子集。

3.4 所提方法的入侵检测模型

如图 2 所示，基于多通道自编码器的深度学习的入侵检测方法主要由多通道自编码器和 1 维 CNN 组成。该模型结合了基于特征检测和异常检测方法的优点，利用正常流量和攻击流量分别训练各自通道的自编码器模型，然后利用 1d CNN 网络提取多通道表示中隐藏的相互关系，从而降低了不平衡数据对模型的影响，极大提升了样本的检测精度，而且由于模型使用了无监督的自编码器和轻量型 1d CNN 网络，有效提高了模型的检测效率。下面给出具体的训练流程：

第一步是利用数据预处理初始化训练样本集；

第二步是利用处理后的样本集对每个通道的自编码器进行训练；

第三步是将不同通道自编码器计算的特征向量进行重构，并作为 1d CNN 的输入；

第四步是利用重构向量训练 CNN，并采用 GA 算法优化 CNN 的超参数；

第五步是利用训练后的模型对样本进行检测，并基于 Softmax 分类器输出分类结果。

在算法 1 中给出了所提方法的伪代码。

算法 1：所提方法的伪代码

1.初始化训练样本集：$\{(x_i, y_i)\}_{i=1}^{N}$ 且 $y_i \in \{attack, normal\}$，单通道样本数据矩阵 X；

2.开始：

3.训练自动编码器：

trainAutoencoder(X^n)→z_n

trainAutoencoder(X^a)→z_a

4. 利用两个自动编码器 z_n 和 z_a 计算重构向量；

5. *for* $x \in X$ *do*

构建多通道样本 $x_i' = \left[x_i, \hat{x}_i^n, \hat{x}_i^a\right]^T$；

end

6.训练基于 GA 优化的 1 维 *CNN*：

基于 GA 优化 CNN 超参数；

训练模型 train1DCNN(X')→*model*

7. 返回结果

4 实验与结果分析

为了评估所提入侵检测方法的有效性，所提方法在 *Python* 3.7 中使用 Keras 2.3 库和 *TensorFlow* 进行测试，并将测试结果与 DNN[1]、AIDA[10]、CNN-1D[13]、Gray-scale[14]、WnD[15]和 MDPCA-DBN[16]等现有的深度学习入侵检测方法进行对比。对于测试数据集，使用 Hyperopt 库中的树状结构 Parzen 估计器算法进行超参数优化，其中使用 20%的训练集作为验证集。而且测试数据使用最小-最大缩放器进行缩放。表 1 给出了实现最佳验证损失的参数配置。

表 1　实验参数配置

Table 1　Experimental parameter configuration

	自编码器	分类器
批量大小	$\{2^5,2^6,2^7,2^8,2^9\}$	$\{2^5,2^6,2^7,2^8,2^9\}$
学习率	[0.0001,0.01]	[0.0001,0.01]
dropout 系数	[0,1]	[0,1]

自动编码器被定义为三层。这些层包括 KDDCUP99 和 UNSW-NB15 数据集中的 40×10×40 神经元和 CICIDS2017 中的 50×10×50 神经元。为了进行数据正则化和防止过拟合，在解码层之前放置一个丢失层。对于自动编码器，选择均方误差作为损失函数。同时，为每个隐藏层选择 *ReLu* 作为激活函数，最后一层则使用线性激活函数。

在分类器方面，该结构由一个一维卷积层和三个完全连通的层组成。该网络以 D×3 大小的样本为输入，预测伯努利概率。输入样本由具有 64 个滤波器的 1D 卷积层转换为尺寸为 D×64 的特征映射，该特征映射分别由尺寸为 320、160 和 2 的全连接层处理。而最后一层的 Softmax 激活函数得到输出概率。为了执行数据正则化，在网络中的每一层后面都有一个丢失层。分类器通过最小化二进制交叉熵作为损失函数来对权重进行优化。

4.1　数据集

本文采用 KDDCUP99[1]、UNSW-NB15[17]和 CICIDS2017[18]三个基准数据集对所提方法进行测试。每个数据集包括标记的训练集(经过处理用于训练入侵检测模型)和测试集(评估经过训练后模型的入侵检测能力)。下面给出 3 个数据集的具体描述。

KDDCUP99 数据集是在 1999 年组织的 KDD 工具竞赛中引入的。这是一个经常用于评估入侵检测系统基准数据集。该数据集包含在军事网络环境中模拟的网络流，并记录为具有 42 个属性的向量，其中训练集包含 4898431 个样本，测试集包含 311027 个样本。为了控制训练阶段的成本，本文采用 10%的训练数据用于训练模型，整个测试集用于评估阶段。此外，整个数据集中的正常流量和攻击流量均不平衡，攻击百分比高于正常流量(训练集中的 80.3%vs19.7%，测试集的 80.5%vs19.5%)。

UNSW-NB15 数据集是由澳大利亚网络安全中心(ACCS)的网络范围实验室中的 IXIA PerfectStorm 工具创建的。该数据集由一个训练集和一个测试集组成，其中包括网络流样本，它们被存储为 43 个属性的向量。训练集中的数据集相当均衡，而测试集中的数据集则存在不平衡，攻击流量的百分比略高于正常流量的百分比(68.1%vs 31.9%)。

CICIDS2017 由加拿大网络安全研究所于 2017 年收集，此数据集包含正常流量和最新的攻击流量。它还包括使用带有标记的流量特征抓取工具 CICFlowMeter 执行的网络流量分析结果。数据集中每个网络流样本都具有 79 个属性。在本文实验中，通过分层随机抽样的方式选择原始日志中的 80%的正常流量和 20%的攻击流量创建训练集和测试集，其中训练集和测试集分别包含 10 万和 90 万个样本。表 2 中给出了三个数据集的相关参数。

表 2　三个数据集的相关参数

Table 2　Relevant parameters of three data sets

数据集		属性	总样本	正常样本占比	攻击样本占比
KDDCUP99	训练	42	494021	19.7%	80.3%
	测试		311029	19.5%	80.5%
UNSW-NB15	训练	43	82332	44.9%	55.1%
	测试		175341	31.9%	68.1%
CICIDS2017	训练	79	100000	80%	20%

	测试		900000	80%	20%

4.2 评估指标

通过分析训练的入侵检测模型的准确率和 *F*-分数来衡量所提出方法的整体性能，这些度量值可以简单地从混淆表中获得。其中，准确率是正确标记流量的比率，定义为：

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

式中：*TP* 和 *TN* 分别表示为正确预测流量为正常和攻击类型的样本数，*FP* 和 *FN* 分别表示为错误预测流量为正常和攻击类型的样本数，*TP+TN+FP+FN* 为总样本数。测试结果的准确率越高，算法正确预测流量类型的性能越高。

F 分数是精度和召回率的调和平均值，其中精度 *p* 衡量入侵检测系统仅识别攻击的能力，召回率 *r* 可以看作是系统发现所有攻击的能力，其定义为：

$$p = \frac{TP}{TP + FP} \tag{8}$$

$$r = \frac{TP}{TP + FN} \tag{9}$$

$$F = \frac{2 \times r \times p}{r + p} \tag{10}$$

F 值越高，算法所达到的精确性和召回率之间的平衡越好。

4.3 结果分析

所提算法采用 4 个实验来验证性能的优越性。第一个实验是消融研究，通过设置神经网络(*neural network，NN*)、人工神经网络(*artificial neural network，ANN*)、CNN 和人工卷积神经网络(*artificial convolution neural network，ACNN*)四种网络结构来验证本文多通道深度学习算法的有效性；第二个实验是鲁棒性，通过在不平衡数据集中进行测试，验证所提算法的稳健性；第三个实验是探讨自编码器 z_n 和 z_a 重建样本的过程，分析多通道自动编码器的有效性；第四个实验是对比研究，通过与其他几种入侵检测算法结果对比验证所提算法的优势。

4.3.1 消融研究

首先给出 NN、ANN、CNN 和 ACNN 四种网络结构的架构：

(1) NN 网络模型由一个输入层和图 2 体系结构的最后 4 个层组成(1 个 Flatten 层，2 个 *FC* 层和一个 Softmax 层)，输入样本为 $x_i \in R^D$，即 $X^{(D)} \rightarrow$ *Flatten*(320)→*FC*(160)→*FC*(2)→*Softmax*。

(2) ANN 网络的架构与 NN 相同，但输入样本采用两个自编码器的组合信息 $x_i \oplus \hat{x}_i^n \oplus \hat{x}_i^a \in R^{3D}$，架构为：$X^{(3D)} \rightarrow$ *Flatten*(320)→*FC*(160)→*FC*(2)→*Softmax*。

(3)CNN 网络的架构是在 NN 中添加一维卷积层，架构为：$X^{(D)}$→*Conv1D*(64)→*Flatten*(320)→*FC*(160) → *FC*(2)→*Softmax*。

(4)ACNN 网络的架构与 CNN 相同，但是输入数据采用自编码器的组合信息，架构为：$X^{(3D)} \rightarrow$ *Conv1D*(64)→*Flatten*(320)→*FC*(160)→*FC*(2)→*Softmax*。

图 3 和图 4 给出了所提算法与 NN，ANN，CNN 和 ACNN 四个网络在准确性和 *F*-分数的测试结果。从图中可以看出，所提算法在准确性和 F-分数指标上明显优于所有基线，从而证实了组合自动编码器，1D 卷积和多通道输入的有效性，能够提高入侵检测任务的准确性。特别要注意的是，与卷积解耦的自动编码器不能保证性能的整体提高。另一方面，撇开自动编码器，卷积密集层通常可以提高入侵检测的准确性。在任何情况下，将卷积应用于自动编码器丰富的数据时，通常都可以实现较高的精度和 F 分数。通过分析发现，卷积导致的模型优越性取决于在多个通道(而不是通过级联构建的单个通道)上

计算卷积的能力，并在原始变量及其基于自动编码器的对应变量中寻找特征。

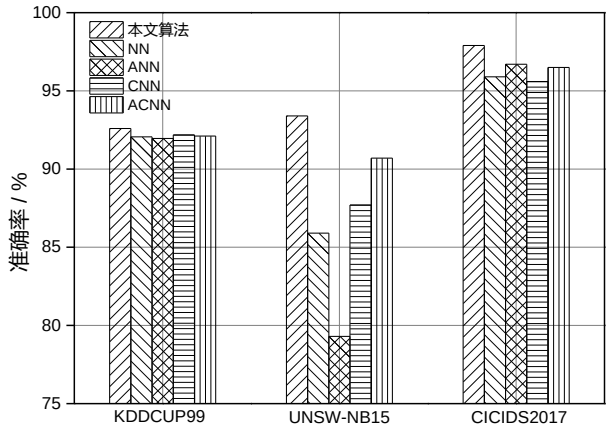


Fig.3 Accuracy of different network models in three datasets

图 3 不同网络模型在 3 个数据集的准确率结果

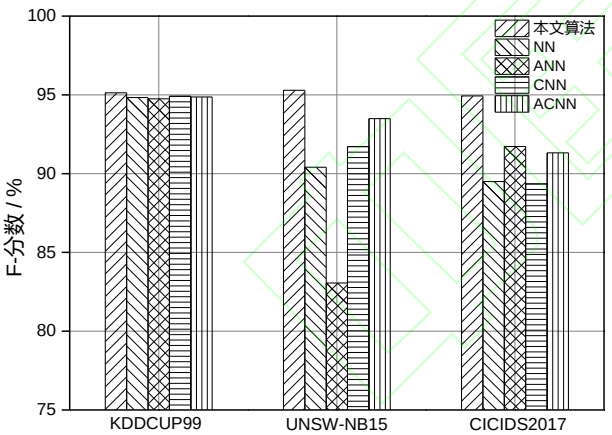


Fig.4 F-score of different network models in three datasets

图 4 不同网络模型在 3 个数据集的 F-分数结果

此外，本文还对比了几个网络估计的参数个数，如表 3 所示。从表中可以看到，所提算法的较高准确度通常是以估计参数的数量较多为代价的。同时，攻击样本占比越小，所需参数越多。

表 3 不同网络的估计参数

Table 3 Estimated parameters of different networks

数据集	所提方法	NN	ANN	CNN	ACNN
KDDCUP99	891920	65120	91360	891810	2571170
UNSW-NB15	912449	65445	92334	912293	2632610
CICIDS2017	1649697	76961	126887	1649573	4844465

4.3.2 鲁棒性

第二个实验主要是用于分析所提方法在解决不平衡数据问题时的鲁棒性。 对于这个实验的分析数据，本文采用 CICIDS2017 数据集，该数据集基于现实世界网络的情景收集到的不平衡数据，其中包括正常流量占 80%，攻击流量占 20%。为了验证所提方法对不平衡数据的鲁棒性，将数据集中的正常流量和攻击流量进行组合，获得攻击流量占总样本的 100%、75%、50%、25%和 5%五个样本子集用于测试。

图 5 给出了所提算法与 NN，ANN，CNN 和 ACNN 四个网络在五个样本子集中的 *F*-分数测试结果。从图中可以看出，但是本文方法对应的 *F*-分数下降幅度最小，在所有方法中 F-分数仍是最高，从而说明所提方法适用于不平衡数据的入侵检测。

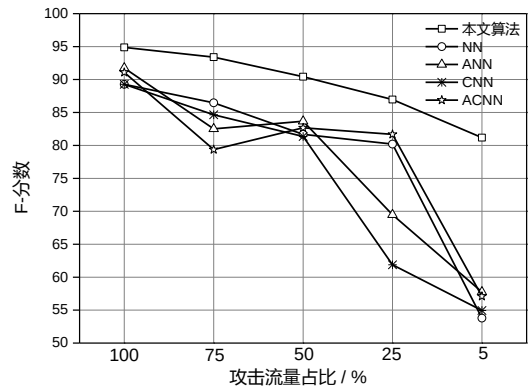
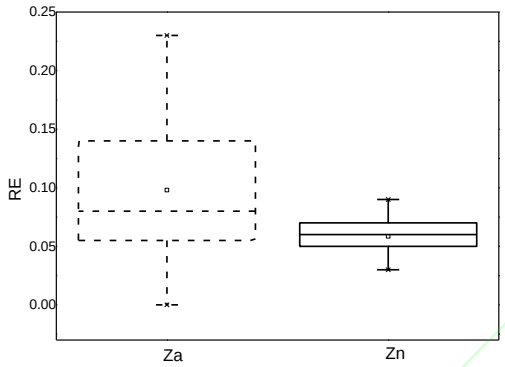


Fig.5 F-score of different network models in unbalanced data

图 5 不同网络模型在不平衡数据的 F-分数结果

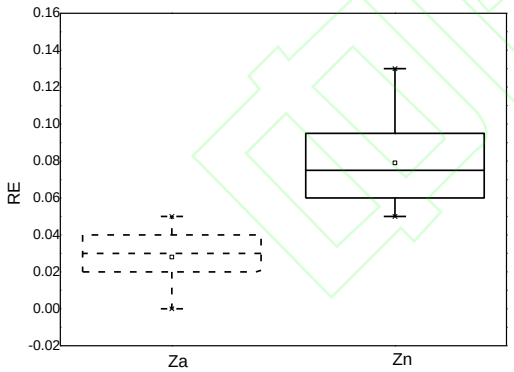
4.3.3　多通道自编码器分析

第三个实验是对多通过自编码器进行分析，探讨自编码器 z_n 和 z_a 准确地从正常和攻击两类样本中重建样本的过程。图 6 显示了当 UNSW-NB15 数据集中使用自动编码器 z_n 和 z_a 重构正常流 x 和攻击流 $\hat{x}$ 时的重构误差箱形图，计算方式为 $RE=\|x-\hat{x}\|^2$ 。



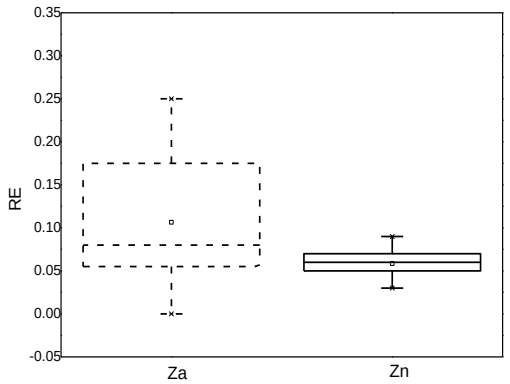
(a) Normal flow of training set

(a)训练集正常流量



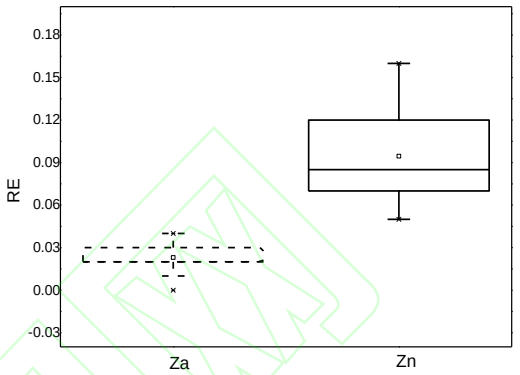
(b)Training set attack traffic

(b) 训练集攻击流量



(c)Normal flow of test set

(c)测试集正常流量



(d)Test set attack traffic

(d) 测试集攻击流量

Fig.6　Reconstruction error analysis

图 6　重构误差分析

从图中可以看出，自动编码器 Z_n 在重建正常样本时比重建攻击样本更准确，同时 Z_a 则具有相反的表现。从而证明了利用多通道自编码器将正常样本和攻击样本分开训练能够引入有助于区分这两个类别的信息。此外，从图中还观察到，所提方法可以更好地受益于两个类别中基于自动编码器的样本数量。

4.3.4　对比研究

第四个实验给出了所提方法与 DNN、AIDA、CNN-1D、Gray-scale、WnD 和 MDPCA-DBN 等几种入侵检测算法结果对比。

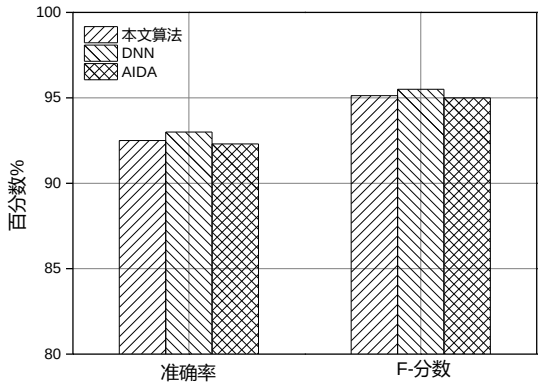


Fig.7 Comparison of different algorithms in KDDCUP99

图 7 不同算法在 KDDCUP99 数据集的对比结果

图 7-图 9 给出了不同入侵检测算法在 3 个数据集上的测试结果对比。从图中可以看出，所提算法在 3 个数据集上测试的准确率和 F-分数结果整体上优于对比方法，充分说明本文方法的有效性和优越性。此外，从图中观察到唯一的例外是在 KDDCUP99 数据集上，所提算法的结果处于次优解，DNN 模型具有最佳的准确率和 F-分数。这是因为 DNN 通过一个深度神经网络以及文本表示方法来学习入侵检测模型，然后在系统调用中捕获上下文和与序列相关的信息，该模型通过优化过程寻找网络的最佳参数和最佳拓扑。因此，在 KDDCUP99 数据集上 DNN 的较高精度可以归因于文本表示方法，以及所确定体系结构的拓扑结构和参数设置。

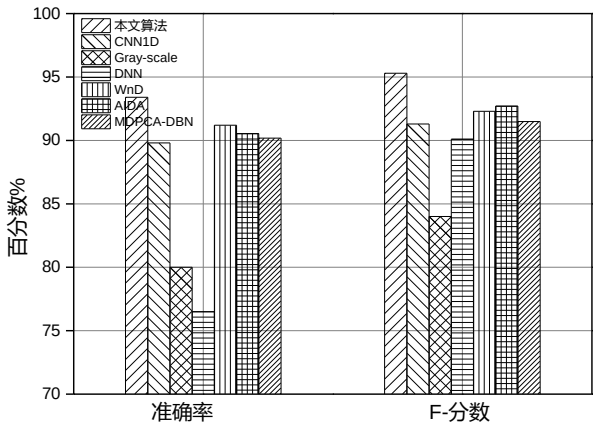


Fig.8 Comparison of different algorithms in UNSW-NB15

图 8 不同算法在 UNSW-NB15 数据集的对比结果

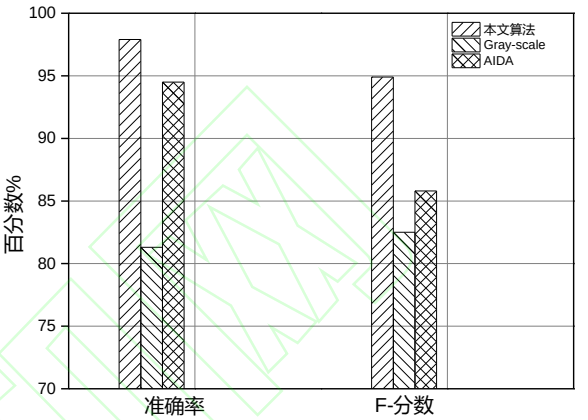


Fig.9 Comparison of different algorithms in CICIDS2017

图 9 不同算法在 CICIDS2017 数据集的对比结果

5 结语

本文提出了一种基于多通道自编码器深度学习的入侵检测方法，用于解决当前入侵检测方法中存在的准确率低、误报率高的问题。所提方法采用无监督的多通道特征学习和有监督的跨通道特征依赖有效地结合的方式构建检测模型：首先在无监督阶段，采用分别采用正常流量和攻击流量训练两个自编码器，原始样本与两个编码器重构出的新特征向量共同组成多通道特征向量表示；然后将多通道特征向量作为 1 维 CNN 网络的输入，用于学习通道之间可能的依赖关系。同时，为了保持模型的灵活性，本文采用遗传算法优化一维 CNN 模型的拓扑结构，以提高入侵检测时的预测性能。实验结果表明，所

提方法能够有效提高入侵检测系统的准确率，比其他方法的测试结果更佳。

References:

[1] Vinayakumar R, Alazab M, Soman K P, et al. Deep Learning Approach for Intelligent Intrusion Detection System[J]. IEEE Access, 2019, 7: 41525-41550.

[2] Naseer S, Saleem Y, Khalid S, et al. Enhanced Network Anomaly Detection Based on Deep Neural Networks[J]. IEEE Access, 2018, 6: 48231-48246.

[3] Hassan M M, Gumaei A, Alsanad A, et al. A hybrid deep learning model for efficient intrusion detection in big data environment[J]. Information Sciences, 2020, 513: 386-396.

[4] 王毅, 冯小年, 钱铁云,等. 基于CNN和LSTM深度网络的伪装用户入侵检测[J]. 计算机科学与探索，2018, 12(4): 575-585.
Wang Yi, Feng Xiaonian, Qian Tieyun, et al. Disguised user intrusion detection based on CNN and LSTM deep network[J]. Journal of Computer Science and Exploration, 2018, 12(4): 575-585.

[5] Vijayanand R, Devaraj D. A novel feature selection method using whale optimization algorithm and genetic operators for intrusion detection system in wireless mesh network[J]. IEEE Access, 2020, 8: 56847-56854.

[6] Labonne M, Olivereau A, Polve B, et al. A Cascade-structured Meta-Specialists Approach for Neural Network-based Intrusion Detection[C]//2019 16th IEEE Consumer Communications & Networking Conference(CCNC). Las Vegas, USA: IEEE, 2019: 1-6.

[7] Lin W H, Lin H C, Wang P, et al. Using convolutional neural networks to network intrusion detection for cyber threats[C]//2018 IEEE International Conference on Applied System Innovation (ICASI). IEEE, 2018.

[8] Almiani M, Abughazleh A, Alrahayfeh A, et al. Deep recurrent neural network for IoT intrusion detection system[J]. Simulation Modelling Practice and Theory, 2020, 101: 1-20.

[9] 张思聪, 谢晓尧, 徐洋. 基于 dCNN 的入侵检测方法[J]. 清华大学学报(自然科学版), 2019, 59(1): 46-54.
Zhang Sicong, Xie Xiaoyao, Xu Yang. Intrusion detection method based on dCNN[J]. Journal of Tsinghua University (Natural Science Edition), 2019, 59(1): 46-54.

[10] Alqatf M, Lasheng Y, Alhabib M, et al. Deep Learning Approach Combining Sparse Autoencoder With SVM for Network Intrusion Detection[J]. IEEE Access, 2018, 6: 52843-52856.

[11] Andresini G, Appice A, Mauro N D, et al. Exploiting the Auto-Encoder Residual Error for Intrusion Detection[C]// 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). Stockholm, Sweden: IEEE, 2019: 281-290,

[12] Mirza A H, Cosan S. Computer network intrusion detection using sequential LSTM Neural Networks autoen-coders[C]//2018 26th Signal Processing and Communica-tions Applications Conference (SIU). Izmir, Turkey: IEEE, 2018: 1-4.

[13] Lopez-Martin, Manuel, et al. Shallow neural network with kernel approximation for prediction problems in highly demanding data networks[J]. Expert Systems with Applications, 2019, 124: 196-208.

[14] Kim T, Suh S C, Kim H J, et al. An Encoding Technique for CNN-based Network Anomaly Detection[C]//2018 International Conference on Big Data(Big Data). Seattle, USA: IEEE, 2018: 2960-2965.

[15] Yan Jiaqi, Jin Dong, Lee C W, et al. A Comparative Study of Off-Line Deep Learning Based Network Intrusion

Detection[C]//2018 10th International Conference on Ubiquitous and Future Networks (ICUFN). Prague, Republic: IEEE, 2018: 299-304.

[16] Yang Yanqing, Zheng Kangfeng, Wu Chunhua, et al. Building an Effective Intrusion Detection System Using the Modified Density Peak Clustering Algorithm and Deep Belief Networks[J]. Applied Sciences, 2019, 9(2): 1-25.

[17] 张蕾，钱峰，赵姝,等. 利用变分自编码器进行网络表示学习[J]. 计算机科学与探索, 2019, 13(10): 1733-1744.

Zhang Lei, Qian Feng, Zhao Shu, etc. Using Variational Autoencoders for Network Representation Learning[J]. Journal of Computer Science and Exploration, 2019, 13(10): 1733-1744.

[18] Harbola S, Coors V. One dimensional convolutional neural network architectures for wind prediction[J]. Energy Conversion and Management, 2019, 195: 70-75.

YANG Jie was born in 1976, male (Yao nationality), from Yongzhou Hunan, a postgraduate of computer science at Central South University, professor of Hunan University of Science and Engineering, His main research interests are network security and artificial intelligence.
杨杰（1976—），男（瑶族），湖南永州人，中南大学计算机硕士研究生，湖南科技学院教授，主要研究方向为网络安全，人工智能。

TANG Yachun was born in 1980, male, from Yongzhou Hunan, Software Engineering Master of Hunan University, Experimenter of Hunan University of science and Engineering, His main research interests are network security and artificial intelligence.
唐亚纯（1980—），男，湖南永州人，湖南大学软件工程硕士研究生，湖南科技学院实验师，主要研究方向为网络安全，人工智能。

TAN Daojun was born in 1975. He received the master degree. He is a Senior Experimentalist at Hunan University of Science and Engineering. His research interests include artificial intelligence, deep learning et al.
谭道军（1975—），男，湖北黄冈人，华中师范大学教育技术学硕士研究生，湖南科技学院高级实验师。主要研究方向为机器学习与深度学习等。

LIU Xiaobing was born in 1989, male, from Yongzhou, Hunan Province, master's degree in control engineering at Guangxi University, lecturer of Hunan University of Science and Engineering, His main research direction is network security and artificial intelligence.
刘小兵（1989—），男，湖南永州人，广西大学控制工程专业硕士研究生，湖南科技学院讲师，主要研究方向为人工智能、网络与信息安全。