文章编号：2095-4980(2024)12-1394-06

基于深度学习与遗传算法的IoT环境多层次威胁溯源方法

曹新立[1]，阎　峻[2]，孙　领[3]，丁晓玲[3]

(1.国网新源控股有限公司 检修分公司，北京 100032；2.国网新源控股有限公司，北京 100032；
3.福建厦门抽水蓄能有限公司，福建 厦门 361100)

摘　要：常规物联网(IoT)环境的多层次威胁溯源主要采用网络数据关系分类实现，忽略了常规数据与威胁数据之间的相似性，导致溯源结果误报数较多。对此，提出基于深度学习与遗传算法的物联网环境多层次威胁溯源方法。建立深度学习神经网络对IoT环境威胁数据进行识别，添加批量标准化操作将常规数据与威胁数据分离，提取多层次威胁数据特征，应用遗传算法得到最优个体，实现威胁数据初始节点的溯源定位。实验结果表明，应用所提方法得出溯源结果误报数较少，溯源结果较为准确，满足IoT环境安全维护的现实需求。

关键词：物联网环境；物联网多层次威胁；威胁溯源；深度学习；遗传算法

中图分类号：TP393.08　　**文献标志码：**A　　**doi：**10.11805/TKYDA2023285

A multi-level threats traceability method for IoT environment based on deep learning and genetic algorithm

CAO Xinli[1]，YAN Jun[2]，SUN Ling[3]，DING Xiaoling[3]

(1.Maintenance Branch，State Grid Xinyuan Holding Co.，Ltd.，Beijing 100032，China；2.State Grid Xinyuan Holdings Co.，Ltd.，Beijing 100032，China；3.Xiamen Fujian Pumped Storage Co.，Ltd.，Xiamen Fujian 361100，China)

Abstract：The multi-level threats attribution in conventional Internet of Things(IoT) environments is mainly achieved through the classification of network data relationships, which overlooks the similarity between conventional data and threat data, leading to a large number of false positives in the attribution results. In response to this, a multi-level threats attribution method for IoT environments based on deep learning and genetic algorithms is proposed. A deep learning neural network is established to identify threat data in the IoT environment, and batch normalization operations are added to separate conventional data from threat data, extracting features of multi-level threats data. Genetic algorithms are applied to obtain the optimal individual, achieving the initial node attribution and positioning of threat data. Experimental results show that the attribution results obtained using the proposed method have fewer false positives and are more accurate, meeting the practical needs for the security maintenance of IoT environments.

Keywords：IoT environment；multi-level threats to the IoT；traceability of threats；deep learning；genetic algorithm

对物联网遭受的网络攻击威胁进行溯源成为物联网安全保证的重要课题，通过对网络攻击的溯源、对网络上的恶意行为进行有效取证，为类似恶意行为带来有效的威慑，从而进一步提高物联网运行过程的安全[1-2]。因此，对网络攻击进行溯源是一项十分关键的应用技术。

文献[3]针对海量数据下的网络攻击溯源问题进行研究，对网络攻击的本体进行分析，以此建立攻击事件的本体模型；结合单一日志下的网络攻击解析其流量数据类型，构建不同数据之间的日志条目关系。基于这一关系提取网络攻击数据中的线索数据。引入图模型理念，在构建的本体模型中嵌入自动学习算法，在优化后的模型中，对网络攻击数据的关键线索特征进行分析与提取，通过对线索特征进行分类，实现网络攻击的溯源过程，

收稿日期：2023-09-28；**修回日期：**2023-12-01

基金项目：国网新源控股有限公司重点资助项目(SGXYXM00JHJS1900051)

但该方法普适性较差。文献[4]针对网络攻击的潜伏期较长的问题进行溯源研究，采用监督学习的方法对网络攻击的多阶段数据进行分析，解析网络攻击的完整过程，以此来组建网络关系图。在该关系图中，对网络数据的日志连接关系进行保护，保障数据的隐私性。引入Louvain算法对关系图进行优化，建立图卷积神经网络，解决状态爆炸问题，并完成对网络数据关系的分类，实现网络攻击的溯源过程，但该方法溯源准确率较低。文献[5]针对网络攻击犯罪事件中的网络取证问题进行研究，建立基于软件定义网络的网络攻击防御模型，在该模型中对网络攻击数据进行模拟，对攻击源的定位精确度不断优化。建立以取证能力为评估指标的评估体系，利用该评估体系对网络攻击防御模型进行迭代，修正网络攻击阻断手法，将成功阻断的手法与网络攻击的数据类型进行对应，分类识别出网络攻击的类型，实现网络攻击的溯源过程，但该方法溯源效率较低。

上述文献提出的网络攻击溯源技术无法满足IoT运行安全保障的现实需求，本文针对IoT环境的多层次威胁，提出一种基于深度学习与遗传算法的IoT环境多层次威胁溯源方法。该方法通过结合深度学习和遗传算法两种技术，实现威胁数据的定位溯源过程。

1　物联网环境多层次威胁溯源方法设计

1.1　识别物联网环境威胁数据

针对IoT在运行过程中存在大量的运行数据以及对威胁数据的识别问题，本文建立威胁数据识别的多层神经网络，采用深度学习的方法对威胁数据进行识别。输入层和输出层均为物联网环境数据传导的网络层，均设定为初始值[6]。本文主要对深度学习多层神经网络的隐藏层进行研究，在物联网的运行数据导入隐藏层后，对隐藏层内所有数据的平均值进行计算：

$$\bar{i}=\frac{1}{m}\sum_{i=1}^{m}x_i \tag{1}$$

式中：$\bar{i}$为隐藏层数据均值；m为隐藏层内的数据量；x_i为隐藏层数据的特征值。

基于式(1)，对隐藏层数据的方差进行计算[7-9]：

$$\sigma^2=\frac{1}{m}\sum_{i=1}^{m}(x_i-\bar{i})^2 \tag{2}$$

基于该方差，构建3×3的深度学习卷积核，如图1所示[10]。建立以数据方差为内核的深度学习卷积核，在该卷积核中，对隐藏层数据进行标准化处理：

$$\hat{i}=\frac{x_i-\bar{i}}{\sqrt{\sigma^2+\varepsilon}} \tag{3}$$

式中：$\hat{i}$为标准化后的数据；ε为卷积核参数。

−1	−1	1
−2	0	2
−1	0	1

Fig.1 Deep learning convolutional kernel
图1 深度学习卷积核

在该深度学习卷积核中加入批量标准化的操作，使3个隐藏层中的物联网数据保持相同的分布[11]。在此基础上，对IoT环境威胁数据进行识别：

$$i_{\mathrm{w}}=f\oplus D(\hat{i})+b \tag{4}$$

式中：i_{w}为物联网威胁数据；f为数据维度；$D(\cdot)$为下采样；b为加性偏置值。通过式(4)的与或运算，将IoT环境中的威胁数据识别出来。根据该识别结果，将IoT中的常规运行数据与威胁数据进行分离，完成基于深度学习的物联网环境威胁数据识别过程。

1.2　提取多层次威胁数据特征

基于识别出来的物联网环境威胁数据，对多层次的威胁数据特征进行提取。对物联网多层次的威胁数据带有某一特征的概率进行表征：

$$P_Q=1-\sum_{i_{\mathrm{w}}=1}^{n}Q \tag{5}$$

式中：P_Q为威胁数据携带某特征的概率；Q为威胁数据的标记信息；n为识别出的威胁数量。

根据计算出的概率，对威胁数据特征的期望分布进行分析[12]：

$$E(Q)=\sum_{i_w=1}^{n}\frac{v}{v-i_w+1} \tag{6}$$

式中：$E(Q)$为对应标记信息的期望分布；v为服从参数。

对威胁数据与数据特征之间的线性关系进行表征[13]：

$$r=\frac{E_{Qi_w}}{\sqrt{E_Q E_{i_w}}} \tag{7}$$

式中：r为威胁数据与其特征的线性关系；E_{Qi_w}为威胁数据与标记信息之间的离均差乘积和；E_Q为标记信息期望分布的离均差平方和；E_{i_w}为威胁数据期望分布的离均差平方和。

根据该线性关系，对物联网多层次的威胁数据特征进行提取：

$$H_i=\sum_{i_w-1}^{s}2^r+k(255-r)^s \tag{8}$$

式中：H_i为提取的威胁数据特征；k为特征提取系数；s为物联网层级。

通过上述步骤，完成物联网环境多层次威胁数据特征的提取过程。

1.3 溯源定位威胁初始节点

根据提取到的物联网环境多层次威胁数据特征，对威胁数据的初始节点进行溯源定位，本文采用遗传算法完成这一步骤。物联网环境中包含若干个网络节点，对威胁数据入侵物联网的初始节点进行定位，实现威胁溯源过程。设定遗传算法的框架，分别为[14]：

Step1：初始化种群；

Step2：适应度计算；

Step3：执行基本操作；

Step4：迭代优化，输出最优个体。

在Step1中，将提取到的威胁数据特征作为个体进行种群的初始化，在物联网环境的多层次架构中随机生成初始种群。完成种群初始化后，对种群的适应度进行计算[15]：

$$F=\sum_{H=1}^{g}Y \tag{9}$$

式中：F为适应度函数；Y为锚点位置；g为权重系数；H为威胁个体的特征量。

在step3中，执行的基本操作包括种群个体的交叉与变异。通过交叉操作，将威胁数据特征个体进行两两配对，并随机产生一个在[0,1]区间内的随机数；基于该随机数，对两两配对的交叉位置进行定位，获得威胁数据的交叉节点。对交叉过程中种群个体的变异概率进行计算：

$$L=\frac{\alpha(F_{\max}-a)}{F_{\max}-\overline{F}} \tag{10}$$

式中：L为种群个体变异概率；α为常数；$F_{\max}$为种群最大适应度值；$\overline{F}$为种群平均适应度值；a为个体适应度值。

将式(10)的计算结果与上述步骤生成的随机数进行对比，判定其是否大于该个体的变异概率。如“是”，则执行变异操作；反之，则不对该个体执行变异操作。

不断循环迭代上述步骤，直至种群个体不再出现改良，输出最优个体。

根据输出的遗传算法最优个体，对威胁数据的初始节点进行定位溯源：

$$G=(F_{\max}+e)^2-R^2 \tag{11}$$

式中：G为威胁数据的溯源节点；e为最优个体值；R为物联网层次节点空间布局。

通过上述步骤，得到威胁数据的初始节点定位信息，完成物联网环境多层次威胁溯源方法的设计。

2　实验

2.1 实验准备

对本文所提的基于深度学习与遗传算法的物联网环境多层次威胁溯源方法的可行性进行测试，并通过对比分析其应用有效性。

搭建实验环境，如图2所示，包括主控计算机和物联网架构外设2个主体结构；同时按照表1和表2所示的配置搭建本次溯源实验的实验环境。

对本次实验所应用的威胁数据进行设置。本次实验采用ISCX–Bot数据集，该数据集混合了IoT数据集以及ISCX 2012 1DS 数据集，符合真实的物联网环境威胁情况，该数据集中的攻击数据类型及数量如表3所示。本次实验共包括10 330条攻击数据，采用该数据对物联网环境进行攻击，分析不同溯源方法的有效性。

完成上述准备后，开展本次物联网环境多层次威胁溯源实验。

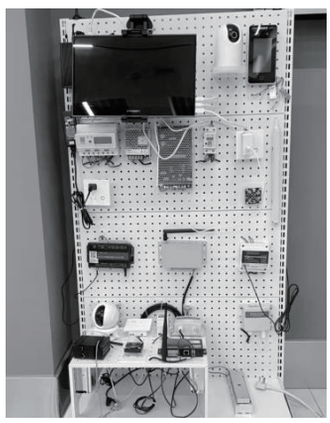

Fig.2 Experimental environment
图2 实验环境

表1 物联网外设通信参数设定

Table1 Setting of communication parameters for IoT peripherals

item	parameter
frequency range	GSM & LTE
modulate	orthogonal frequency division multiple access BPSK,OPSK
channel bandwidth/kHz	180
signal bandwidth/kHz	3.75
data speed/kbps	200
access mechanism	LTE
data frame maximum load/ bytes	1 600
Reto reception sensitivity/dBm	–130

表2 主控计算机配置

Table2 Main control computer configuration

configure	parameter
CPU	Inter Pentium G2020 @ 2.90 GHz
Operating System(OS)	Windows 10
Motherboard M	CM6731_CM6431_CM6331
internal storage	16 GB
primary hard drive	500 GB;700 rad/min
video card	Nvidia GeForce 505

2.2 物联网环境多层次威胁溯源

采用实验数据集中的攻击数据对物联网环境进行威胁测试，所提方法对该威胁的溯源结果如图3所示。由图3可知，本文所提方法对物联网环境的3个层次受到的威胁进行了溯源，得到了不同端口受到威胁的数据路径。与本次测试导入的攻击数据真实路径结果进行对比可知，本文所提方法的溯源结果与真实路径结果高度一致。从这一实验结果可以初步判断，本文提出的溯源方法具备可行性。

表3 攻击数据类型及数量

Table3 Types and quantity of attack data

data type	quantity
probe	415
Dos	1 526
R2L	2 354
U2R	1 654
Data	126
DDoS	4 255

2.3 结果评价指标

IoT环境划分为不同的通信路径，对不同路径的威胁溯源结果的单条路径的收敛数目进行统计，计算攻击数据属于该路径的概率：

$$P_l=\sum_{l=1}^{3}p\left(\frac{x\times l}{d}\right) \tag{12}$$

式中：P_l为属于对应路径的概率；l为物联网环境的数据传输路径；x为实验测试攻击数据；d为该路径的长度。

通过式(12)获得某一物联网路径接收到威胁数据包时的归属概率。基于此，对不同方法的溯源有效性进行判断，统计出单条路径的收敛数目，并根据该数目对溯源结果的误报数进行计算：

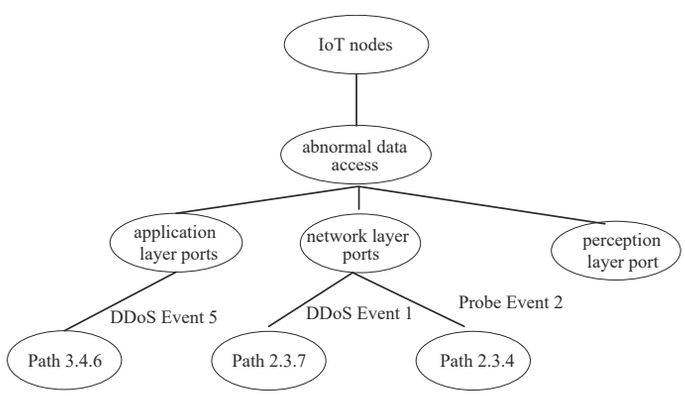

Fig.3 Traceability of multi–level threats in the IoT environment
图3 物联网环境多层次威胁溯源

$$M=1-\left(1-\frac{1}{2^d}\right)^{l\times 3} \tag{13}$$

式中M为溯源结果的误报数。

基于该计算结果，得到不同溯源方法的误报数。误报数数值越低，表明对应方法得到的威胁溯源结果误报数越少，有效性越高，能够对不同路径的威胁数据进行精准溯源。

2.4 结果分析与讨论

分别采用文献[3-5]提出的网络攻击溯源方法作为本次实验的对比方法，将几种不同溯源方法得到的结果进行横向对比分析，测试不同方法的有效性。

为保证本次实验结果的科学性，不同溯源方法导入的攻击数据保持完全一致；同时，为减少实验误差，多次进行同组实验，取多次结果的平均值作为最终的实验结果进行有效性分析。经过实验，得到不同方法下的物联网环境多层次威胁溯源实验结果。统计不同方法在物联网多层次环境中的收敛数目，如表4所示。根据表4，得到不同方法对IoT环境多层次威胁溯源实验的误报数结果，如图4所示。

表4 不同方法的收敛数目

Table4 Convergence numbers of different methods

IoT path	number of convergence /bars			
	the method of this paper	reference [3] method	reference [4] method	reference [5] method
application Layer Path 1 #	2 624	1 837	2 014	1 645
application Layer Path 2 #	1 356	1 154	954	1 277
network layer path 1 #	1 566	1 340	1 292	1 067
network layer path 2 #	464	287	315	193
sense layer path 1 #	1 364	1 057	1 296	1 104
sense layer path 2 #	2 269	1 846	2 154	2 095

根据表4和图4可知，本文所提方法在单条路径中的收敛数目较高，由此得出本文所提方法在进行威胁溯源时的误报数较少，表明本文所提方法的溯源准确率较高。而其他3种溯源方法的实验结果中，统计所得的单条路径收敛数目与本文所提方法有着较大的差距，计算得出的误报数也比较大，溯源准确率较低。

从这一实验结果可以看出，本文所提方法对物联网环境的多层次威胁的溯源结果较为准确，能够有效为物联网的安全保障工作提供助力。

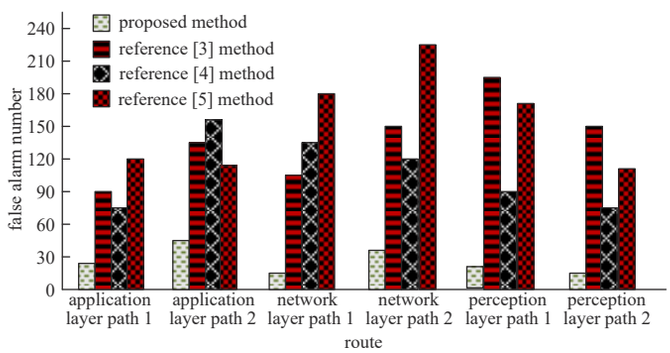

Fig.4 False alarm count results of different traceability methods

图4 不同溯源方法的误报数结果

3 结论

物联网作为一种应用范围较为广阔的网络技术，其安全问题是领域内共同关注的问题。在多层次的IoT环境中，对遭受的攻击威胁进行精准溯源成为一个关键技术。对此，本文提出了一种基于深度学习与遗传算法的物联网环境多层次威胁溯源方法。经过实验可知，本文所提方法统计所得的单条路径收敛数目较高，误报数较少，能够有效为物联网现实应用过程中的安全保障工作提供科学指导，进一步推动物联网技术的应用与发展，具备广阔的应用前景。

参考文献：

[1] 周全兴,李秋贤,王振龙,等. 基于博弈论的无线传感网络节点攻防优化[J]. 太赫兹科学与电子信息学报, 2022,20(2):181-187,193. (ZHOU Quanxing,LI Qiuxian,WANG Zhenlong,et al. Attack and defense optimization of wireless sensor network nodes based on game theory[J]. Journal of Terahertz Science and Electronic Information Technology, 2022,20(2):181-187,193.) doi:10.11805/TKYDA2020659.

[2] 温祥彬,郑媛. 应用于物联网云安全可信度检测的算法仿真[J]. 计算机仿真, 2022,39(5):225-228. (WEN Xiangbin,ZHENG Yuan. Algorithm simulation applied to cloud security credibility detection of internet of things[J]. Computer Simulation, 2022, 39(5):225-228.) doi:10.3969/j.issn.1006-9348.2022.05.045.

[3] 黄克振,连一峰,冯登国,等. 一种基于图模型的网络攻击溯源方法[J]. 软件学报, 2022,33(2):683-698. (HUANG Kezhen, LIAN Yifeng,FENG Dengguo,et al. Method of cyber attack attribution based on graph model[J]. Journal of Software, 2022,33(2): 683-698.) doi:10.13328/j.cnki.jos.006314.

[4] 李腾,乔伟,张嘉伟,等. 隐私保护的基于图卷积神经网络的攻击溯源方法[J]. 计算机研究与发展, 2021,58(5):1006-1020. (LI Teng, QIAO Wei, ZHANG Jiawei, et al. Privacy-preserving network attack provenance based on graph convolutional neural network[J]. Journal of Computer Research and Development, 2021, 58(5): 1006-1020.) doi: 10.7544/issn1000-1239.2021.20200942.

[5] 刘雪花,丁丽萍,郑涛,等. 面向网络取证的网络攻击追踪溯源技术分析[J]. 软件学报, 2021,32(1):194-217. (LIU Xuehua, DING Liping, ZHENG Tao, et al. Analysis of cyber attack traceback techniques from the perspective of network forensics[J]. Journal of Software, 2021,32(1):194-217.) doi:10.13328/j.cnki.jos.006105.

[6] 曾凡锋,田雨丝,王景中. 物联网中节点捕获攻击早期检测方法研究[J]. 计算机仿真, 2022,39(8):477-481. (ZENG Fanfeng, TIAN Yusi, WANG Jingzhong. Research on early detection method of node capture attack in Internet of things[J]. Computer Simulation, 2022,39(8):477-481.) doi:10.3969/j.issn.1006-9348.2022.08.091.

[7] 张琳,李焕洲,张健,等. 基于物联网集成防御机制的诱饵路径优化算法[J]. 计算机应用研究, 2021,38(11):3433-3438. (ZHANG Lin, LI Huanzhou, ZHANG Jian, et al. Decoy path optimization algorithm based on integrated defense mechanism of internet of things[J]. Application Research of Computers, 2021, 38(11): 3433-3438.) doi: 10.19734/j. issn. 1001-3695.2021.04.0134.

[8] 李群,董佳涵,关志涛,等. 一种基于聚类分类的物联网恶意攻击检测方法[J]. 信息网络安全, 2021,21(8):82-90. (LI Qun, DONG Jiahan, GUAN Zhitao, et al. A clustering and classification-based malicious attack detection method for internet of things[J]. Netinfo Security, 2021,21(8):82-90.) doi:10.3969/j.issn.1671-1122.2021.08.010.

[9] 刘新,黄缘缘,刘子昂,等. IoTGuardEye:一种面向物联网服务的Web攻击检测方法[J]. 计算机科学, 2021,48(2):324-329. (LIU Xin, HUANG Yuanyuan, LIU Ziang, et al. IoTGuardEye: a Web attack detection method for IoT services[J]. Computer Science, 2021,48(2):324-329.) doi:10.11896/jsjkx.200800030.

[10] 王艳芬,丁宇,陈瑞瑞,等. 基于深度学习的物联网设备通信低功耗资源分配算法[J]. 南京邮电大学学报(自然科学版), 2022, 42(1): 6-12. (WANG Yanfen, DING Yu, CHEN Ruirui, et al. A low-power resource allocation algorithm for IoT device communication based on deep learning[J]. Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2022,42(1):6-12.) doi:10.14132/j.cnki.1673-5439.2022.01.002.

[11] 李贝贝,宋佳芮,杜卿芸,等. DRL-IDS:基于深度强化学习的工业物联网入侵检测系统[J]. 计算机科学, 2021,48(7):47-54. (LI Beibei,SONG Jiarui,DU Qingyun,et al. DRL-IDS:deep reinforcement learning based intrusion detection system for industrial internet of things[J]. Computer Science, 2021,48(7):47-54.) doi:10.11896/jsjkx.210400021.

[12] 王宏刚,孙明月,简燕红,等. 电力物联网中基于深度学习的无蜂窝接入点选择算法[J]. 现代电力, 2021,38(5):529-534. (WANG Honggang,SUN Mingyue,JIAN Yanhong,et al. Deep learning based cell-free access point selection algorithm in power internet of things[J]. Modern Electric Power, 2021,38(5):529-534.) doi:10.19725/j.cnki.1007-2322.2020.0365.

[13] 莫桂江. 蚁群-遗传算法的无线传感器网络路径优化[J]. 微电子学与计算机, 2011,28(9):139-142. (MO Guijiang. Wireless sensor network path optimization based on ant colony-genetic algorithm[J]. Microelectronics & Computer, 2011, 28(9): 139-142.)

[14] 李政宇,李练兵,芮莹莹. 基于改进遗传算法优化自联想神经网络的风机故障诊断[J]. 计算机应用与软件, 2022,39(6):297-302,328. (LI Zhengyu, LI Lianbing, RUI Yingying. Fault diagnosis of wind turbine based on auto-associative neural network optimized by improved genetic algorithm[J]. Computer Applications and Software, 2022,39(6):297-302,328.) doi:10.3969/j.issn.1000-386x.2022.06.044.

[15] 张泽,王瑛,闫孟达,等. 基于启发式遗传算法的近距空中支援超网络结构优化[J]. 兵器装备工程学报, 2022,43(6):121-127. (ZHANG Ze, WANG Ying, YAN Mengda, et al. Optimization of close air support super-network structure based on heuristic genetic algorithm[J]. Journal of Ordnance Equipment Engineering, 2022,43(6):121-127.) doi:10.11809/bqzbgcxb2022.06.020.

作者简介：

曹新立(1989-)，男，硕士，工程师，主要研究方向为企业信息安全.email:Kikoone1@163.com.

阎　峻(1985-)，男，硕士，高级工程师，主要研究方向为数字化电站建设.

孙　领(1988-)，男，本科，工程师，主要研究方向为自动化控制.

丁晓玲(1995-)，女，本科，助理工程师，主要研究方向为企业信息安全.