基于图卷积网络的CAN总线异常检测模型

刘恒汭，张淼
（北京邮电大学网络空间安全学院，北京 100876）

摘要：目前，智能网联汽车产品已经在全球范围内得到了广泛应用，汽车的攻击面也随之扩大，车内总线网络的安全受到威胁。现有的基于深度学习技术的CAN总线网络异常检测模型，通常选取CAN总线报文的部分特征，削弱了报文数据段和整个报文序列所携带的信息。并且，这些模型无法兼顾CAN总线丰富的攻击类型，过高的计算复杂度也使得实时异常检测具有困难。为解决这些问题，本文提出了一种图级别的基于图卷积网络GCN的CAN总线异常检测模型，通过CAN消息区间图结构表征CAN数据信息，利用图卷积层在区间图上学习节点及结构信息，并对各层输出进行加权聚合，池化获得整图的嵌入表示，进行异常检测。实验结果表明，在四种类型的CAN总线异常数据集上，所提出的模型均具有较高的检测精度。

关键词：网络空间安全；车内CAN总线；异常检测；图卷积网络
中图分类号：TP309.2

CAN Bus Anomaly Detection Model Based on GCN

LIU Hengrui, ZHANG Miao
(School of Cyberspace Security, Beijing University of Posts and Telecommunications. Beijing 100876)

Abstract: At present, the intelligent connected automobile have been widely used in the world, the attack surface of the automobile also expands, and the security of the bus network in the car is threatened. The existing CAN bus network anomaly detection models based on deep learning technology usually select some features of CAN bus messages, which has weakened the information carried by the CAN data and message sequence. Moreover, these models cannot take the rich attack types of CAN bus into account, and the high computational complexity makes real-time anomaly detection difficult. To solve these problems, a graph-level CAN bus anomaly detection model based on graph convolutional network (GCN) is proposed in this paper. CAN data information is characterized by CAN message interval graph structure, and the graph convolution layer is used to learn node and structure information on the interval graph, the output of each layer is weighted and aggregated, and the embedded representation of the whole graph is obtained by global pooling operation for anomaly detection. The experimental results show that the proposed model has high detection accuracy on four types of CAN bus anomaly datasets.

Key words: Cyberspace Security; CAN bus; Anomaly detection; Graph convolution network;

0 引言

近年来，智能网联汽车的发展引起了汽车行业的巨大变革，智能网联汽车通过搭载功能复杂的软硬件实现了丰富的功能，提供了更加舒适的驾驶体验。然而，智能网联汽车的发展引入了车辆网络开放的新场景，为CAN总线的应用提出了新问题。攻击者可以利用各类传感器、Wi-Fi以及OBD等等设备侵入车内CAN总线，能够窃听甚至注入恶意数据，威胁行车安全。因此，对车载总线的防护直接关系到驾驶员的人身和财产安全，具有重要意义。

针对车载CAN总线网络的研究方向中，异常检测系统能够以较低的成本部署在现有的

作者简介：刘恒汭（1997-），男，硕士研究生，主要研究方向数据安全
通信联系人：张淼（1980-），男，副教授、硕导，主要研究方向数据安全. E-mail: zhangmiao@bupt.edu.cn

车内网络，在既定的 CAN 总线网络中保证信息的安全，因此更加适配车载网络场景下的安全研究。已有的车载总线异常检测模型在面对 CAN 总线丰富的攻击类型存在着一定的局限性，对于伪造和重放等攻击类型的检测效果有所下降，并且模型训练与计算的复杂度较高，实时性不佳。

为全面考虑 CAN 总线的数据信息，对现有 CAN 总线异常检测方案作出改进，本文提出了一种基于图卷积网络 GCN 的 CAN 总线异常检测模型。模型将 CAN 总线报文序列划分为连续区间，并构建对应的为 CAN 消息区间图结构，通过节点特征记录报文传输的数据信息，边连接刻画报文在序列中的关系；使用多层图神经网络层对区间图进行学习，并对各层的输出加权聚合，生成节点的特征向量表示；通过图池化层生成区间图的嵌入表示，用于图级别的异常检测。

1 相关工作

CAN 总线数据具有一些明显的特征，可以被基于统计的异常检测方法所利用。CAN 总线协议使用标识符 ID 来表征报文与 ECU 的对应关系，某一 ID 的报文通常由一个固定的 ECU 发送出来；汽车运行过程中产生着大量的状态物理量，这些物理量需要被实时传输，以供驾驶员参考和 ECU 作出相应动作，因此有非常多 ID 承载的报文被 ECU 以固定的周期向总线中发送；CAN 总线协议的广播特性、仲裁机制、错误处理机制都为研究人员提供了可供参考的数据特征。

现有的针对 CAN 总线异常检测的研究可以分为基于数据特征统计和基于机器学习分类的方法。有研究文献[1-5]引入信息熵、发送频率、ID 转换模式等特征构建异常检测模型，但面对重放或模仿总线正常频率的攻击表现不佳，并且无法检测对报文数据段的篡改。随着机器学习在异常检测任务中的应用和发展，很多研究[6-10]将深度神经网络、长短期记忆网络（LSTM）、生成式对抗网络（GAN）等技术应用于 CAN 总线的异常检测领域，取得了较好的检测效果。Li[11]等提出了一种基于迁移学习技术和卷积神经网络（CNN）的异常检测模型，将 CAN 报文序列转换为 RGB 图片，集成多个在图像识别领域表现优秀的卷积神经网络模型（如 VGG16、Xception 等）进行报文的异常识别。Zhang[12]等提出了一种基于联邦学习和图神经网络（GNN）的异常检测模型，以 CAN 报文区间构造图结构，用两阶段的分类模块实现异常检测和攻击类型分类。这些模型在检测其针对的攻击类型时具有良好的检测精度，但是很难对多种攻击类型均作出有效防护。为更好检测注入与伪造攻击，现有的研究在同时考虑 CAN 消息频率和内容的基础上，需要分别考虑多个 CAN ID，对每个 ID 单独收集数据并训练机器学习模型，数据的单独收集会引入较高延迟，多个机器学习模型也增加了模型计算的复杂度，不能适应 CAN 总线异常检测任务高实时性的要求。

综上可见，CAN 总线的异常检测重点关注 CAN 报文的数据信息和传输特征，选取有效的特征信息并建模对 CAN 总线的异常检测效果具有重要意义。将 CAN 总线数据转换为图结构，利用图神经网络技术能够全面对 CAN 总线报文的数据特征以及报文之间的相关关系建模，使模型充分利用 CAN 报文序列传递的信息，因此本文选择基于图神经网络的技术实现异常检测模型。

2 基于 GCN 的 CAN 总线异常检测模型

本文提出了一种基于图卷积网络 GCN 的 CAN 总线异常检测模型。模型将 CAN 总线报文序列划分为连续区间，并构建对应的为 CAN 消息区间图结构，通过节点特征记录报文传输的数据信息，边连接刻画报文在序列中的关系；使用多层图神经网络层对区间图进行学习，并对各层的输出加权聚合，生成节点的特征向量表示；通过图池化层生成区间图的嵌入表示，用于图级别的异常检测。图神经网络模型的结构如图 1 所示：

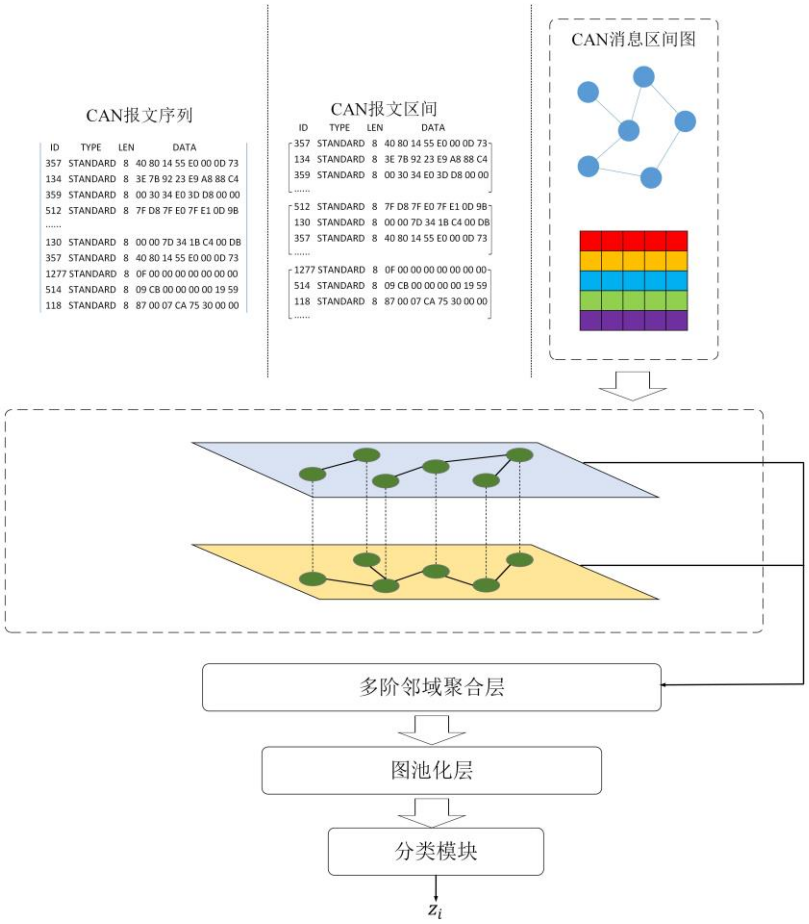

图 1 CAN 总线异常检测模型

Fig. 1 Anomaly Detection Model on CAN Bus Data

2.1 CAN 消息区间图转化

模型将连续时间段采样的报文区间转化为对应的 CAN 消息区间图样本，作为后续模型的输入。具体构造步骤如下：

步骤 1：划分 CAN 消息区间。将 CAN 消息的原始数据按照发送的时间顺序划分为若干个连续的 CAN 消息区间，每个区间内包含的 CAN 报文数量相同。

步骤 2：生成图节点。为步骤 1 划分出的每个 CAN 消息区间对应生成一个图样本，图样本中的节点表示 CAN 消息区间中的报文，每条报文与其所在区间的图样本中唯一一个节点对应。节点的特征向量来自报文所承载的信息，节点的特征向量包含 10 个维度，分别为 ID、时间戳以及 8 个数据段，其中，8 个数据段由报文传输的数据段部分以字节为单位划分

得到。

步骤 3：生成边连接。报文区间中每两条相邻的报文所对应的两个节点之间生成一条无向边连接。在报文序列中 ID 相同，且位置最为临近的前后两条报文所对应的两个节点之间生成一条无向边连接。

步骤 4：对区间图标注。为区间图样本添加标签，以标识正常、异常图数据。区间内报文全部为正常报文，则标签值为 0，代表正常图数据；区间内存在异常报文，则标签值为 1，代表异常图数据。

由于 CAN 消息区间图单个样本的规模较小，节点和边数量有限，每次训练迭代输入单个图样本会造成计算资源浪费，影响训练效率。算法通过构建图样本的 mini-batch 来实现多个图样本训练的并行化。选取多个图样本，构造一个包含这些独立子图的巨大图，其中，各个子图的邻接矩阵以对角化方式堆叠，子图的节点特征矩阵进行拼接操作。

2.2 图神经网络层

将 CAN 消息区间图G_T作为输入，目标是得到区间图的图嵌入向量。图神经网络层用来融合不同的节点特征，将节点特征映射到低维空间，生成节点嵌入；图池化操作来聚合不同的节点嵌入，最终输出图向量，用以表示子图G_T的图嵌入；最后用分类器对图嵌入转换为对图的预测。

图神经网络层让节点特征沿着图的结构传播聚合，将图节点及其结构信息映射到低维空间，生成节点的特征向量表示，即节点的嵌入，聚合邻居节点的特征信息并且更新输出目标节点的特征向量。

在本文的研究中应用图卷积神经网络（GCN）[13]作为图神经网络层，将区间图G_T的邻接矩阵和图节点的特征矩阵作为图神经网络层的输入，叠加第 l 层图神经网络层能够使中心节点聚合其第 l 阶内邻域的邻居信息。经过多层图神经网络层的迭代，图中的节点特征表示聚合了其多层邻域节点的信息。

2.3 多阶邻域聚合层

由于 CAN 消息区间图的结构较为简单，节点和边数量有限，在经过多层图神经网络层的聚合后，节点聚合了多阶邻域的信息，彼此的特征表示会趋于一致，节点特征表示的多样性降低后，节点之间的可区分性变弱。因此，为避免模型性能因图神经网络层的层数增加而下降，模型在形成图嵌入前，增加节点多阶邻域表示的聚合层，将前段多个图神经网络层的输出进行加权聚合，形成节点的最终表示：

$$Z_i = \sigma\left(\sum_{k=1}^{N} q_k W_k h_i^{(k)}\right) \tag{1}$$

其中，$h_i^{(k)}$为第 k 层图神经网络层输出的节点 i 的特征表示，q_k为可学习的权重参数，W_k为非线性变换矩阵，将各层输出节点表示转换为相同维度，σ 为激活函数。

2.4 图池化层

为了获得图级别的向量表示，图池化层将图神经网络层输出的所有节点特征读出，将节点特征聚合得到一个统一的图嵌入，作为区间图的图嵌入。

图池化层使用全局均值池化（Global Average Pooling）[14]方法，将图中所有节点特征向

量按各个维度求平均值，得到区间图的图嵌入。

2.5 分类模块

区间图G_T通过图池化层输出的图嵌入可以用Z_T来表示。分类模块以 Softmax 函数作为分类器，将图嵌入Z_T输入到 Softmax 函数来预测图的类别标签y_T。

2.6 目标函数

训练过程中，模型通过计算区间图G_T的标签值y_T与预测的区间图标签概率分布$\tilde{y}_T$之间的损失函数，结合反向传播（Back Propagation）和梯度下降（Gradient Descent）法进行模型参数优化。模型使用交叉熵（Cross Entropy）作为损失函数。

3 实验结果与分析

3.1 数据集介绍

实验使用韩国 HCRL 实验室提供的 Car-Hacking 数据集[8]，该数据集由一辆正常行驶中的现代起亚轿车采集得到，包括正常报文、拒绝服务（DoS）攻击、模糊（Fuzzy）攻击、针对挡位（gear）和转速（RPM）的欺骗攻击的数据集。报文数据属性包括：时间戳、CAN ID、DLC、DATA[0~7]、Flag 标签。Flag 标签值为 R 代表正常报文，T 代表异常报文。

表 1 数据集总览

Tab. 1 Overview of CAN Bus Dataset

数据集类型	报文总数	正常报文数量	异常报文数量
正常报文	988871	988871	0
拒绝服务(DoS)攻击	3665771	3078250	587521
模糊（Fuzzy）攻击	3838860	3347013	491847
转速（RPM）欺骗攻击	4443142	3845890	597252
挡位（gear）欺骗攻击	4621702	3966805	654897

每个数据集内包括采集 30~40 分钟的 CAN 总线报文数据，包含 300 次注入攻击，每次攻击持续 3~5 秒。在 DoS 攻击场景中，攻击者每隔 0.3ms 注入一条 ID 为‘0000’的报文；在模糊攻击场景中，攻击者每隔 0.5ms 注入一条 ID 与数据完全随机的报文；在欺骗攻击场景中，攻击者每隔 1ms 注入与转速/挡位相关的特定 ID 的报文。

3.2 对比模型及参数设置

为验证本章所提出的模型的有效性，将其与图神经网络的基准模型共同在 CAN 总线异常数据集上进行了对比实验。这些基准模型分别是：

GCN[13]：它是一种适用于直推式学习任务，使用全图节点信息进行训练的谱域图卷积网络；

GraphSAGE[15]：它通过采样提取节点的邻域特征并进行聚合，来生成目标节点的嵌入表示；

GAT[16]：它是结合注意力机制的图神经网络，通过图注意力机制学习节点及其邻居的重要性。

在实验中，所有模型的批处理样本数设置为 64，节点嵌入维度设置为 64，初始学习率为 0.001，使用 Adam 优化器，图神经网络层数为 2，dropout 设为 0.5，统一迭代 150 次；；CAN 消息区间图的节点数设为 40。

本节用于评价模型表现的指标为准确率（Accuracy）、召回率（Recall）和 F1-Score。

3.3 结果分析

对本文所提出的模型及对比模型在四组异常数据集上的检测结果进行展示和分析，实验结果如表 2 所示：

表 2 实验结果

Tab. 2 Experimental Results

数据集	评价指标	GCN	GraphSAGE	GAT	GNN_CAN
DoS	Accuracy	99.44	99.44	99.67	99.64
	Recall	98.92	99.44	99.39	**99.62**
	F1	99.46	99.72	99.69	**99.81**
Fuzzy	Accuracy	98.40	99.00	99.33	**99.40**
	Recall	95.77	96.63	97.75	**98.88**
	F1	97.84	98.27	98.86	**99.44**
RPM	Accuracy	98.33	97.78	97.78	**99.13**
	Recall	**98.92**	97.70	97.75	98.70
	F1	99.13	98.84	98.65	**99.29**
gear	Accuracy	96.67	96.67	99.00	**99.30**
	Recall	95.69	96.67	96.88	**98.78**
	F1	97.80	98.30	98.41	**99.40**

表 2 展示了本章提出的针对 CAN 总线异常检测任务提出的模型和其他基于图神经网络的基准模型在不同攻击类型数据集上的检测结果。对于考察的大多数的指标，可以观察到本章所提出的模型的检测效果优于其他基准模型，并且在不同攻击类型的数据集上表现稳定。在模糊攻击数据集上基准模型的召回率指标有所下降，其潜在原因是模糊攻击使用随机方式生成异常数据，ID 与数据段特征均不固定且无规律，导致基准模型未能准确识别随机生成的少部分异常报文，而本章所提出的模型在模糊攻击数据集的检测仍保持了较高的召回率，能够有效识别随机生成的异常数据；对于拒绝访问攻击数据集，各个模型的表现均很出色；对于针对指定车辆 ECU 的数据集，本章所提出的模型的查全率相对较高，能够准确识别注入的异常报文，其潜在原因是模型考虑了特定 ID 不同报文的的频率与数据特征。

4 结论

本文提出了一种基于图神经网络的 CAN 总线异常检测模型，对 CAN 总线数据进行图级别的异常检测。将 CAN 总线报文序列划分为连续区间，通过构造 CAN 消息区间图结构来融合报文传输的数据信息和序列的结构信息。通过多层图神经网络聚合图中各个节点信息，并对各层图神经网络层生成的输出加权聚合，得到节点的特征表示，通过池化生成区间图的整图嵌入，判断是否存在异常。实验结果验证了模型的有效性，可以检测到绝大多数注入攻击，与其他对比模型相比，模型在不同的异常数据集上都具有较高的检测精度。

[参考文献] (References)

[1] Müter M, Asaj N. Entropy-based anomaly detection for in-vehicle networks[A]. //2011 IEEE Intelligent Vehicles Symposium[C], Baden-Baden: IEEE, 2011:1110-1115.
[2] Wang Q, Lu Z, Qu G. An Entropy Analysis Based Intrusion Detection System for Controller Area Network in Vehicles[A]. // 2018 31st IEEE International System-on-Chip Conference (SOCC), Arlington: IEEE, 2019:90-95
[3] Marchetti M, Stabili D, Guido A, et al. Evaluation of anomaly detection for in-vehicle networks through information theoretic algorithms[A]. // 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)[C], Bologna: IEEE, 2016:1–6.
[4] Taylor A, Japkowicz N, Leblanc S. Frequency-based anomaly detection for the automotive CAN bus[A]. // 2015 World Congress on Industrial Control Systems Security (WCICSS)[C], London: IEEE, 2015:45–49.
[5] Marchetti M, Stabili D. Anomaly detection of CAN bus messages through analysis of ID sequences[A]. // 2017 IEEE Intelligent Vehicles Symposium (IV)[C], Los Angeles: IEEE, 2017:1577–1583.
[6] Kang M J, Kang J W. A Novel Intrusion Detection Method Using Deep Neural Network for In-Vehicle Network Security[A]. // 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)[C], Nanjing: IEEE, 2016:1-5.
[7] Taylor A, Leblanc S, Japkowicz N . Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks[A]. // IEEE International Conference on Data Science & Advanced Analytics[C], Montreal: IEEE, 2016:130-139.
[8] Seo E, Song H M, Kim H K. GIDS: GAN based intrusion detection system for in-vehicle network[A]. // 2018 16th Annual Conference on Privacy, Security and Trust (PST)[C], Belfast: IEEE, 2018:1–6.
[9] Groza B, Murvay P S. Efficient Intrusion Detection With Bloom Filtering in Controller Area Networks[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(4):1037-1051.
[10] Amato F, Coppolino L, Mercaldo F, et al. CAN-Bus Attack Detection With Deep Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(8):5081-5090.
[11] Yang L, Shami A. A Transfer Learning and Optimized CNN Based Intrusion Detection System for Internet of Vehicles[A]. // ICC 2022 - IEEE International Conference on Communications[C], Seoul: IEEE, 2022:2774-2779.
[12] Zhang H, Zeng K, Lin S, Federated Graph Neural Network for Fast Anomaly Detection in Controller Area Networks[J]. IEEE Transactions on Information Forensics and Security, 2023, 18:1566-1579.
[13] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[A]. // International Conference on Learning Representations[C]. 2016.
[14] Lin M , Chen Q , Yan S. Network In Network[J]. arXiv, arXiv:1312.4400, 2013.
[15] Hamilton W L , Ying R, Leskovec J, et al. Inductive Representation Learning on Large Graphs[A]. // Proceedings of the 31st International Conference on Neural Information Processing Systems[C], Long Beach: Curran Associates Inc, 2017:1025-1035.
[16] Velikovi P, Cucurull G, Casanova A, et al. Graph Attention Networks[A]// International Conference on Learning Representations[C]. 2018.