基于流量特征的未知协议自动化逆向分析方法

李雨晴[1)] 李卓群[1)] 陈 晶[1)] 何 琨[1)] 杜瑞颖[1)]
孙熙平[1)] 吴 聪[2)]

[1)](武汉大学国家网络安全学院 武汉 430040)
[2)](香港大学工学院电气电子工程系 香港 999077)

摘 要 协议逆向分析的核心任务是从未知协议中提取协议格式和语义信息，这是许多协议安全分析技术的基础。由于二进制协议缺乏分隔符，且可读性差，导致其分析复杂度较高，成为协议逆向分析中的重要挑战。现有的协议逆向分析方法通常依赖人工对协议格式的预设假设，这使得其缺乏通用性且自动化水平较低，难以适应复杂且多样化的协议类型。为了解决这些问题，本文提出了一种基于流量特征的未知协议逆向分析方法。该方法通过充分利用流量中多维度特征信息，结合U-Net图像语义分割网络和GRU循环神经网络，从已知协议的网络流量中有效提取协议格式和语义特征，并训练模型以完成对未知协议的逆向分析任务。实验结果表明，本文方法在格式提取准确率上相比现有方法提高了至少19%，在语义推断任务中，平均准确率超过64%，且具备更高的通用性和自动化水平。该技术路径能够显著提升未知协议分析的准确性和效率，尤其在需要对海量网络流量进行快速分析的实际应用场景中，具有重要的应用价值。

关键词 未知协议逆向；流量特征；协议格式提取；协议语义推断；深度学习
中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2025.02984

An Automated Reverse Engineering Method for Unknown Protocols Based on Traffic Features

LI Yu-Qing[1)] LI Zhuo-Qun[1)] CHEN Jing[1)] HE Kun[1)] DU Rui-Ying[1)]
SUN Xi-Ping[1)] WU Cong[2)]

[1)](School of Cyber Science and Engineering, Wuhan University, Wuhan 430040)
[2)](Department of Electrical and Electronic Engineering, College of Engineering, University of Hong Kong, Hong Kong 999077)

Abstract Protocol Reverse Engineering (PRE) is a critical and foundational task in network security, primarily aiming to extract the protocol format and semantic information from unknown communication protocols. This extracted information is essential as it forms the basis for various security analysis techniques, including vulnerability scanning, fuzz testing, intrusion detection, and malicious software behavior analysis. Binary protocols present a significant challenge and complexity during the reverse engineering process due to the lack of delimiters and poor readability, making them a key focus of analysis. Existing methods for PRE typically suffer from several limitations, often relying on manually assumed protocol formats, a dependency that

收稿日期：2025-08-07；在线发布日期：2025-10-16。本课题得到国家自然科学基金（No. 62302343，No. 62472323，No. 62172303）、湖北省重点研发计划（No. 2024BAB018）、武汉市科技成果转化项目（No. 2024030803010172）、山东省重点研发计划基金（No. 2022CXPT055）、武汉市星地融合新一代无线通信产业创新联合实验室（No. 4050902040448）资助。 **李雨晴**，博士，副研究员，主要研究领域为网络安全、分布式系统安全。E-mail：li.yuqing@whu.edu.cn。 **李卓群**，硕士研究生，主要研究领域为网络安全、物联网安全。 **陈 晶**，博士，教授，主要研究领域为网络安全，分布式系统安全。 **何 琨**，博士，副教授，主要研究领域为网络安全，云计算安全。 **杜瑞颖**，博士，教授，主要研究领域为网络安全，隐私保护。 **孙熙平**（通信作者），博士研究生，主要研究领域为移动安全。E-mail：xiping@whu.edu.cn。 **吴 聪**，博士后，主要研究领域为分布式系统安全。

severely restricts their versatility and maintains a low level of automation, rendering them unsuitable for handling the complexity and diversity of protocol types. To address these significant issues, this paper proposes Binary Protocol Reverse Engineering using Neural networks (BPREN), a novel automated PRE method based on traffic features. This approach fully leverages the multidimensional feature information contained within network traffic, strategically combining a U-Net image semantic segmentation network and a Gate Recurrent Unit (GRU) network. The model is robustly trained on the known protocol traffic to effectively and accurately extract both protocol format and semantic features, which are then applied to reverse engineer unknown protocols. The BPREN architecture is built around a sophisticated dual-module design: the protocol format extraction module and the protocol semantic inference module. Specifically, the format extraction module draws inspiration from image processing, treating multiple protocol messages arranged by bit value as a protocol image. It utilizes a carefully modified U-Net network to learn the field boundary features from these known protocol images and subsequently predicts the boundaries in unknown traffic. A probability-based field aggregation algorithm is implemented to reduce false positives and yield a more accurate final protocol format. The semantic inference module follows, using the segmented fields as input for semantic type prediction. It employs a GRU network to extract comprehensive multidimensional features from the protocol fields, considering the field value itself, the same-field context, and the cross-field contextual information. Fine-grained semantics of known protocols are clustered into semantic categories using a Natural Language Processing (NLP) model and k-means clustering. The optimal number of categories is determined automatically by identifying the point of maximum curvature on the curve of the within-cluster sum of squares, which serves as the optimal inflection point for clustering. Furthermore, a contrastive learning objective is employed, coupled with data augmentation, during training to significantly enhance the model's generalization ability to unknown protocols. Extensive experimental results demonstrate the significant effectiveness of the proposed method. The method improves the format extraction accuracy, as measured by the F1 score, by at least 19% compared to existing methods, including Netzob and NEMESYS. In the semantic inference task, when tested on similar protocols, the average accuracy exceeds 64% (combining Acc1 and Acc2). Overall, BPREN offers higher versatility and automation, significantly enhancing the accuracy and efficiency of unknown protocol analysis, which is highly valuable for real-world scenarios that demand rapid analysis of massive network traffic.

Keywords unknown protocol reverse engineering; traffic characteristics; protocol format extraction; protocol semantic inference; deep learning

1 引 言

随着互联网技术的迅猛发展，我们正逐步迈入一个万物互联的时代。据中国互联网络信息中心（CNNIC）发布的第56次《中国互联网络发展状况统计报告》显示[1]，2025年上半年，我国移动互联网接入流量已达到1867亿GB，物联网终端用户数已突破28.31亿。这一发展趋势导致了网络协议数量的激增，然而与HTTP、SMTP等由RFC标准规范的公开协议相比，物联网应用中大量采用的是私有协议，这些协议的格式和语法通常未公开。这种私有协议的未知性给网络监管和安全分析带来了显著挑战。

对于网络安全分析人员而言，准确的协议建模是许多安全分析技术（如漏洞扫描[2-3]、模糊测试[4-6]、入侵检测[7]、恶意软件行为分析[8-9]）的前提。例如，在进行网络协议模糊测试时，必须获取协议的语法、语义及状态信息，从而生成符合特定格式的输入并

指导测试执行。因此，未知协议的网络逆向分析在网络安全领域扮演着至关重要的角色。

网络协议因应用场景的不同而呈现出高度的多样性。以自动驾驶领域为例，CAN总线和FlexRay协议被广泛应用；在工业控制系统中，Modbus和DNP3协议则是常见的选择；此外，许多其他网络应用也大量使用专有协议。传统的人工分析方法不仅效率低、周期长，而且高度依赖分析人员的经验。因此，设计一种通用性强且具备高自动化的协议逆向分析方法显得尤为迫切。近年来，尽管已经有多种网络协议逆向分析方案提出，但现有的协议逆向分析方法在通用性、自动化程度和适应性方面仍存在显著局限性，需要探索新的技术路径来应对不断变化和多样化的协议分析需求。这些现有方案大体上可以分为两类：

第一类方法是基于网络流量的逆向分析。这类方法的输入数据来自协议实体之间的交互，主要通过比对多条协议消息来推断协议规范。该方法不依赖于协议对应的软件信息，仅通过网络流量实例即可提取协议信息。其优点在于运行速度较快、易于自动化，并且具备较强的可扩展性。然而，这类方法对流量样本的完整性和质量极为敏感。如果流量样本不足或特征不明显，最终的分析结果会受到较大影响。此外，目前的流量分析方法大多针对非加密流量，而对密态流量的分析仍面临较大困难。

第二类方法是基于指令执行的逆向分析[10]。这类方法主要通过分析协议实现的应用程序指令来提取协议的格式和语义信息，通常采用污点分析[11]、符号执行[12-13]等软件分析技术。与基于流量的方法相比，基于指令执行的方法对于数据样本的完备性要求较低，理想情况下能够提供较高的分析准确度。然而，在很多应用场景下，研究人员往往无法直接获取协议实体软件或通信设备，即使获得设备，也可能由于芯片防护机制的存在而无法提取固件。此外，二进制程序分析技术本身复杂度较高，效率较低，物联网设备指令集和硬件平台的多样性也在一定程度上限制了此类技术在大规模协议逆向分析中的应用。

针对未知协议自动化逆向分析的挑战，本文提出了一种基于网络流量的深度学习分析方法。由于未知协议的二进制代码难以获取，传统的基于二进制文件的分析方法无法有效应用，因此本文主要依赖网络流量数据进行协议逆向分析。为了解决流量样本的不完整性问题，本文创新性地采用深度学习技术，从已知协议中提取有用特征信息，并将其与未知协议的流量数据相结合，通过深度学习模型进行推断，最终提出并实现了一个高效且通用的未知协议分析模型BPREN。

本研究的核心创新在于针对当前协议逆向分析中存在的两个主要问题：一是流量样本的不完整性，二是协议自动化分析的通用性和效率低下，提出了一种全新的技术路径。通过深度学习，特别是U-Net和GRU网络的结合，本文克服了传统方法在处理非加密流量时所面临的瓶颈。通过对已知协议流量的深度特征学习，结合未知协议流量的推断，本方法在协议格式提取与语义推断的准确性和效率上均得到了显著提升。该技术不仅在实验中表现出较高的准确率，还具备较强的通用性和可扩展性，能够适应不同网络环境和协议类型的自动化逆向分析。

总体而言，本文的贡献可以总结为以下：

（1）提出了一种基于深度学习的二进制未知协议逆向分析方法。该方法能够在协议规范未知的情况下，同时完成协议的格式提取和语义推断，为网络协议的自动化分析提供了一种新的解决方案。

（2）在协议格式提取任务中，采用了图像语义分割网络U-Net，并针对协议格式提取任务对模型进行了优化。同时，通过引入字段聚合机制，提升了模型在不同协议上的通用性和准确性。此外，基于预训练模型，本文方法在处理大量数据时保持了较高的效率。

（3）在协议语义推断任务中，本文利用循环神经网络GRU从多个维度提取协议字段的特征，确保了信息的完整性。通过对已知协议的细粒度语义进行聚合，并优化损失函数，结合数据增强技术，提升了模型的泛化能力，使其能更好地适应未知协议的分析任务。

（4）本方案通过语义类别模板对未知协议的字段语义进行有效推断，即使在数据量较小的情况下，通过使用相似协议进行训练，依然能够达到较高的准确率。

2 相关工作

随着物联网的迅速发展，出现了大量未知的网络协议，对于未知协议的逆向分析研究也日趋增多。目前基于网络流量的协议逆向分析方法一般遵循以下流程[14]，首先对未知网络流量进行预处理，并基于处理数据完成协议的格式提取，然后对提取的字段推断其语义，最后综合格式和语义信息等完成

协议的状态机推理，并进一步形成协议规范。现有的研究既有聚焦于单个任务模块的分析效果，也有关注于整个逆向分析过程的组合优化。下面将对与本文任务相关的研究展开介绍。

2.1 未知协议格式提取研究现状

协议格式提取是网络协议逆向的首要任务，它旨在从相同类型的未知协议消息中识别其结构，本文将目前主要格式提取方案按照其技术路线分为以下几类：

2.1.1 序列对齐方法

序列对齐方法源自生物信息学，通过多序列对齐算法Multiple Sequence Alignment（MSA）从协议信息中提取协议格式，常见的序列对齐算法有Needleman Wunsch（NW）算法和Smith Waterman（SW）算法。Beddoe等[15]最早使用多序列对齐算法来进行网络协议逆向，通过对齐多条协议消息序列来提取协议中的公共字段，之后将公共字段不断地递归合并以提取协议的格式。为了能够让MSA方法适用于更加复杂的语义，Cui等[16]提出了Discoverer方法，该方法首先将序列按照类型划分为token，然后按照规则筛选出Field Distinguisher（FD）字段，并基于FD字段对序列进行递归聚类，最后在同类序列上使用MSA算法，根据token的匹配程度推断出协议格式。Netzob方法[17]在已有的序列对齐算法上进一步改进，在多序列比对的过程中加入了语义信息，并优化了得分矩阵，最终在多种未知协议上取得了很好的效果。Luo等[18]人的工作在序列对齐方法基础上通过与服务器之间的主动通信获取额外高质量信息，但其增加了与服务器之间的交互成本且需要对服务器的持续访问能力。

2.1.2 概率统计方法

概率统计方法利用了消息中不同值的出现频率不同，如果某个候选值频繁出现在协议消息中，那么它就更有可能是协议的关键字字段。如Wang等[19]采用概率统计的方式推断协议格式，通过Kolmogorov-Smirnov（K-S）方法在N-gram算法生成的候选字段中提取协议关键词，并采用最大化原则重构协议头部。Ye等[20]首先通过MSA算法在序列中提取候选词，然后通过协议特性构建约束得到候选词的联合概率分布，最终根据预聚类的结果来计算后验概率，并选择后验概率最大的候选词作为协议的关键词，并基于此关键词对未知协议做出准确的聚类。Liang等[21]观察到工控协议控制字段的特殊性，结合概率分析实现对未知工控协议更加准确的逆向。现有的概率统计方法能比较准确地定位关键词字段，但是在统计的过程中需要依赖于经验设置参数，不能做到完全的自动化。

2.1.3 深度学习方法

深度学习方法主要通过深度学习模型对未知协议进行分类、字段划分或者语义提取。Zhao等[22]使用长短期记忆全卷积神经网络Long Short Term Memory-Fully Connected Network（LSTM-FCN）对未知工业控制协议的格式进行提取，首先对大量公开的工业控制协议按照定义的字段类型进行划分，并将划分好的数据按照时序关系输入到LSTM-FCN网络中进行预训练，然后用训练好的模型对工业控制协议的格式进行推断。PREUNN方法[23]结合多种类型的神经网络来开发协议逆向工具，提出了一系列模块化方法。Zhang等[24]人利用循环神经网络来完成协议关键字的提取，并进一步按关键字进行协议分类，同时在关键字提取和协议聚类之间加入了共享学习层对两部分进行联合优化，减少了串行形式下的错误传播，从而得到更好的关键字提取和协议聚类结果。深度学习方法具有比较好的通用性，而且在预训练的模型下推断过程效率较高，但是由于深度模型的固有特点，其可解释性较差，且对于数据特征提取的依赖性也比较强。

2.2 未知协议语义推断研究现状

协议语义推断旨在推断字段所属的语义信息，帮助研究人员更好地理解协议，相较于格式提取而言，目前针对语义推断的研究较少，按照其实现主要可以分为两类。

2.2.1 基于字段特征的语义推断

不同语义的字段往往在取值上表现出不同的特点，因此基于字段特征的语义分析手段主要通过字段值的特点来确定字段语义。例如Cui等[25]人在RolePlayer中，通过查找已知值的表示来搜索端点地址和参数字段，并进一步结合MSA算法来识别长度、cookie和无关字段。Bermudez等[26]人提出的FieldHunter方法针对不同字段语义提出对应检测方案，通过信息熵和相关性分析等手段对不同字段的语义特点进行识别。在最新的BinaryInferno方法[27]中，作者设计了多种原子检测器，根据字段值的变化特点分别识别浮点数、时间戳、长度等多种字段语义类型，并设计了聚合算法对分析结果进行了合并，减少了字段语义冲突，提升了正确性。目前基于字段特征的语义已经能够对部分特殊语义字段进行高效识别，但对未知字段的覆盖率仍不足。

2.2.2 基于模板匹配的语义推断

基于模板匹配的语义推断需要通过收集字段信息建立语义模板，并进一步利用语义模板来推断未知字段语义。例如Kleber等[28]人首先将消息字段按值向量化，然后计算向量之间的Canberra距离评估字段之间的相似度，最后使用DBSCAN算法根据相似性来将字段聚类为类型模板，通过类型模板来推断未知协议字段语义。Wang等[29]针对工业控制协议进行语义推断，首先分析了典型的工业控制协议字段特征，然后基于字段及其位置特征生成语义匹配模板，之后同样使用模板来推断语义，同时加入了聚类和对齐算法提升准确性。Yang等[30]人针对工业控制协议，将字段语义抽象为特征序列，并基于特征序列模板推断未知协议字段，但只考虑到字段维度特征信息。基于通用模板的匹配方法可以提升字段语义识别的范围，但模板的构建需要准确的特征信息。

3 基于流量特征的未知协议逆向方法

本节主要介绍本文设计的未知协议逆向分析方法BPREN，其主要包含协议格式提取模块和协议语义推断模块，协议格式提取模块首先对未知协议的格式进行提取，即完成未知协议字段的分割，之后协议语义推断模块对分割字段的语义进行推断，分析得到字段所属的语义类别，下面将对两个模块分别进行介绍。

3.1 协议格式提取模块

二进制格式的未知协议，没有明显的字段分隔边界，因此基于分隔符的格式提取方案并不能获得很好的表现。同时由于二进制协议字段之间的差异可能较大，因此基于序列对齐的方法也存在困难，此外由于序列对齐算法的递归特性，其在处理大量未知协议数据时也存在效率缺陷。

目前的协议格式提取方法大多是通过比对多个消息序列来提取协议信息，本文从图像语义分割任务中得到启发，如果把多个协议消息按照比特值在图像上进行排列，通过对图像的分析也可以提取出协议消息之间的特征。因此，在协议格式提取模块中，本文采用了图像语义分割网络U-Net，通过对协议图像的像素级标注来获取图像信息，完成未知协议的格式提取。

3.1.1 总体设计

本节介绍协议格式提取模块的总体设计，模块的整体架构如图1所示，主要包含了协议图像生成，协议格式提取和字段聚合。

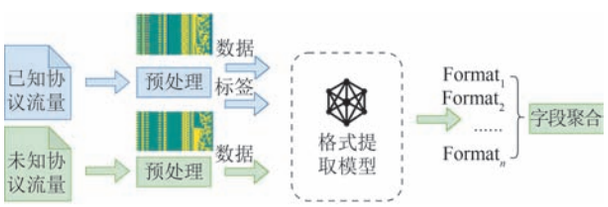

图1 协议格式提取框架

协议图像生成过程对协议样本数据包进行处理，生成特定格式的训练和预测协议图像，若是已知协议，将进一步提取协议分类标签供模型训练。格式提取模型采用图像语义分割网络U-Net，使用已知协议消息图像和边界标签进行训练并对未知协议做出预测。在格式提取模型完成未知协议的字段边界预测后，通过字段聚合算法对多个协议图像的预测结果进行聚合，得到最终的协议格式。下面对各部分具体操作和技术细节做详细阐述。

3.1.2 协议图像生成

协议图像生成过程对获取的网络数据包按协议和会话进行分类，并将分类后的数据拼接生成特定的图像形式。协议图像生成的具体流程如图2所示，主要包含了数据解析、协议聚类、消息分组和图像生成四部分，各部分分别对特定格式数据进行处理。

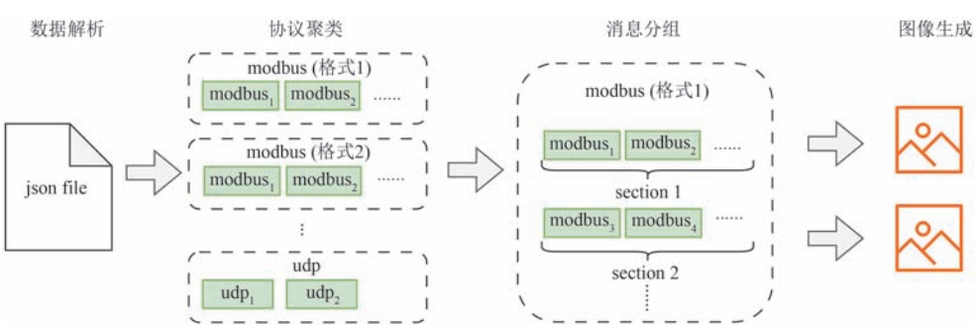

图2 协议图像生成

（1）数据解析：使用协议解析工具tshark对网络流量数据包进行解析，tshark工具可以将pcap数据转化为json文件形式，在json格式文件中，能够获得协议和字段的层次描述，便于后续的处理。

（2）协议聚类：协议聚类过程对解析生成的json格式文件进行处理，首先将消息按照协议封装关系进行分割，保留所需样本协议数据并去除冗余消息。然后通过json文件中的字段偏移量提取出样本协议字段边界并保存为训练标签，字段边界存在的位置被标记为1，其他位置被标记为0。最后样本数据将按照协议和字段边界集合进行聚类。

（3）消息分组：完成协议聚类后的数据可能来自于多个会话，为了更好地提取特征信息，聚类结果将按照会话进行分组。分组的依据来自于解析阶段提取的相关信息，包含IP地址、端口等。同时对于同一组的消息，首先按照数据流进行排序，同一数据流的消息再按照时序关系进行排列。

（4）图像生成：对每一个消息分组内的消息进行拼接，生成协议消息图像。为了统一模型输入尺寸，并兼顾处理效率，本文对协议图像的规格进行了统一设定：协议图像的规格被设置为64，也就是每幅图最多包含64个协议消息，且每条协议消息的长度被限制为32字节（256比特）。对于长度小于32字节的消息，本文使用-1作为填充；对于长度大于32字节的消息，本文采用递归处理机制。具体来说，将截断后的剩余负载数据重新构建为新的消息，并按32字节的规格进行处理，这一策略确保了模型能够分析消息的完整内容。另外，为了最大化利用数据并且保证图像能提取到合理的特征，本文将图像的消息阈值设置为32，具体来说，如果图像包含的消息超过32条，则图像被保留；如果消息数不足32，则图像被丢弃。保留图像的剩余消息位置同样使用-1作为填充。

3.1.3　U-Net格式提取模型

U-Net网络模型是格式提取方法的核心，这里使用的是图像语义分割网络U-Net[31]。传统的U-Net网络的输出是和原图像大小相似的二维图像，而在协议格式提取任务中，最终输出是一维的字段分割结果。为了让U-Net网络适用于二进制协议格式提取的工作，本文对U-Net结构进行了修改，修改后的结构如图3所示，下面将对细节展开介绍。

在改进的U-Net结构中，64条协议消息被拼接成图像作为一组输入，经过U-Net处理得到协议分割边界的预测输出，输入输出的维度可以表示为$R^{n\times 1\times 64\times 256}$，$R^{n\times 1\times 1\times 256}$，其中$n$表示batch的规模大小。

U-Net是一种编码-解码结构，编码器部分通过卷积层和池化层逐级下采样，完成特征提取，解码器部分通过反卷积逐级上采样，完成图像恢复。其中U-Net在编码器和解码器之间还增加了跳跃连接，通过拼接特征提取和上采样部分的特征图来结合高层次和低层次的语义信息，有助于提升图像的精度并保留边缘特征。图3的左半部分表示U-Net的特征提取过程，其中卷积层包含两次卷积运算，卷积核大小为3×3，步长为(1,1)，两次卷积运算增加了输出通道数，同时提取了多特征信息。左半部分的竖向箭头表示卷积层之间的下采样过程，在原始的U-Net结构中，卷积层之间使用池化层进行连接，这里本文采用卷积操作进行了替代，这样能够在减少特征图大小的同时保留更多的特征信息。

本文在上采样过程中将卷积层第二次卷积操作的步长设置为(2,1)，其核心功能在于将相邻两条消息的特征进行融合，降低纵向特征的维度，这个过程旨在让模型专注于区分协议字段的横向边界，从而最终达到一维协议格式输出的需要。同时为保证跳跃连接过程图像拼接尺寸的一致性，本文这里也使用了该卷积层对特征提取部分的特征图进行处理。最终经过U-Net模型处理，二维的协议消息图像中的字段特征会被提取，输出一维的格式提取向量，其中输出比特1的位置代表字段之间的边界。

在训练过程，需要制定损失函数来使模型学习字段特征，完成协议字段划分。本文的损失函数分为两部分，第一部分$\mathcal{L}_{FS}$表示字段分割位置的损失，如公式1所示，其中t_i表示第i比特位置标签值，n表示图像的宽度，$\widehat{y_i}=\mathrm{sigmoid}(y_i)$表示模型的输出的预测值。对于$\mathcal{L}_{FS}$，只有在$t_i=1$的边界位置才有价值，当边界值预测正确($\widehat{y_i}=1$)的时候才不会造成损失。

$$\mathcal{L}_{FS}=\sum_{i=1}^{n}t_i(t_i-\hat{y}_i)^2 \tag{1}$$

第二部分$\mathcal{L}_{NFS}$如公式2所示，$\mathcal{L}_{NFS}$表示非字段分割位置的损失，只有在$t_i=0$的非边界位置才有价值，并且预测为非边界($\widehat{y_i}=0$)时才不会造成损失。

$$\mathcal{L}_{NFS}=\sum_{i=1}^{n}(1-t_i)(t_i-\hat{y}_i)^2 \tag{2}$$

考虑边界损失和非边界损失两部分，总的损失

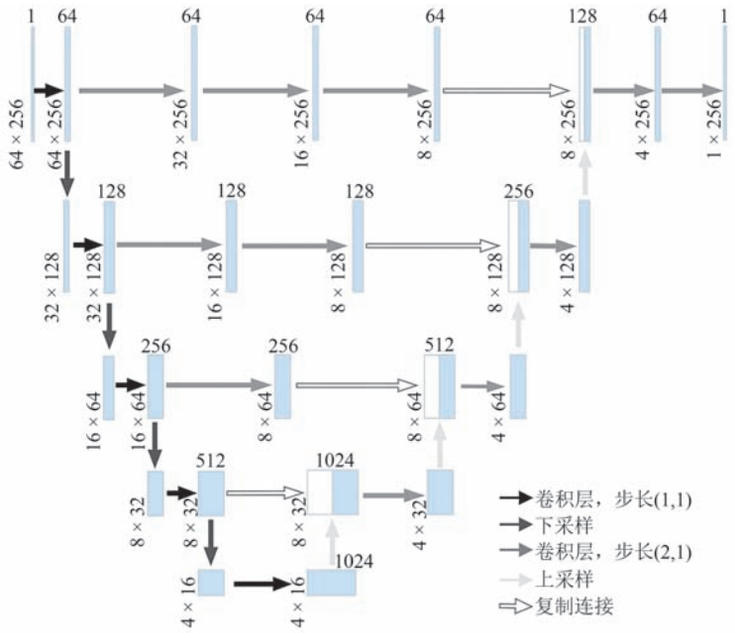

图3 本文所采用的U-Net网络结构

函数如公式3所示，其中λ表示权重系数。在协议字段分割任务中，更多地关心字段分割位置，因此需要给字段边界赋予更多的权重。

$$\mathcal{L}=\lambda\mathcal{L}_{FS}+(1-\lambda)\mathcal{L}_{NFS} \tag{3}$$

3.1.4 字段聚合算法

未知协议的不同消息分组，其图像所预测出的字段边界很可能是不同的，因此为了得到更准确的协议提取格式，需要对不同图像预测得到的字段边界进行聚合。聚合的直观想法是将字段边界进行叠加，输出所有预测结果中出现的边界，这样虽然可以提升对真实边界的覆盖率，但也造成了大量的假阳性情况，增加了研究人员进行边界验证的困难。

本文在这里提出了一种基于概率判断的字段聚合算法，它能够基于预测边界点出现的概率对结果进行聚合，从而降低了简单叠加算法中假阳性出现的可能性。具体过程如算法1所示，首先遍历不同图像所预测的字段边界，保留非重复边界位置，得到位置序列$[pos_1, pos_2, \cdots, pos_n]$。然后对于序列中的每一个位置$pos_i$，计算输出为1(边界存在)的图像个数并记为$N_p$，接下来计算$P_i=N_p/N$，$N$表示所有图像的个数，$P_i$则表示$pos_i$位置输出边界存在的结果所占比例。同时定义一个最小支持度阈值P_t，如果最终$P_i>P_t$，则认为该位置边界存在，反之边界为不存在。P_t的选择基于交叉验证，将训练数据划分为k个子集，对于每个候选的阈值(例如从0.1到0.9，步长0.1)，在$k-1$个子集上训练模型，并在剩余的一个子集上评估其性能，记录每一个阈值对应的平均性能指标，选择具有最高性能指标的阈值作为最终的最小支持度阈值。

算法1： 字段聚合算法

输入：不同协议图像预测结果$pred[N]$，N为图像数量，阈值P_t

输出：协议边界的最终预测向量$result$

1：初始化输出数组$result$，定义集合$temp$

2：FOR 每一个图像预测结果$pred[i]$，$i\in\{1,2,\cdots,N\}$ DO

3： FOR $pred[i]$中每一个预测位置pos DO

4： $temp$集合保留非重复位置pos

5： END FOR

6：END FOR

7：FOR $temp$集合中的位置元素pos DO

8： $N_p\leftarrow 0$

9： FOR 每个图像预测结果$pred[i]$，$i\in\{1,2,\cdots,N\}$ DO

10： IF pos在$pred[i]$中 THEN

11： $N_p=N_p+1$

12： END IF

13： END FOR

14： IF $N_p/N>P_t$ THEN

15： 将pos位置加入最终预测向量$result$中

16： END IF

17：END FOR

3.2 协议语义推断模块

协议语义推断在协议字段分割的基础上进行，

它使用划分好的协议字段作为输入，对协议字段的语义类型做出预测。本文的协议语义推断模块借鉴了模板匹配方法，首先学习已知协议字段特征生成语义类别模板，然后基于语义类别模板对未知协议的语义进行相似性的推断。模板生成需要字段信息，本文基于循环神经网络GRU对字段特征进行提取，和现有方案相比，本文提取方案不仅考虑到字段值本身，还加入了字段的上下文信息提升了特征提取的准确性。同时本文使用NLP模型提取语义信息并使用聚类算法对提取向量进行了聚类，减少了细粒度语义对推断的影响。

3.2.1　总体设计

图4展示了语义推断模块的工作流程，分为训练和推断两部分。训练阶段首先需要对已知协议数据进行预处理，提取已知协议字段特征和相应的语义，接下来按照语义相似度将细粒度语义聚合为语义类别，最后将字段提取特征和语义类型聚类之间进行映射，训练语义推断模型。

语义推断阶段使用训练模型对未知协议语义进行推断，首先要使用格式提取模块对未知协议的字段边界进行划分，然后使用训练好的语义推断模型

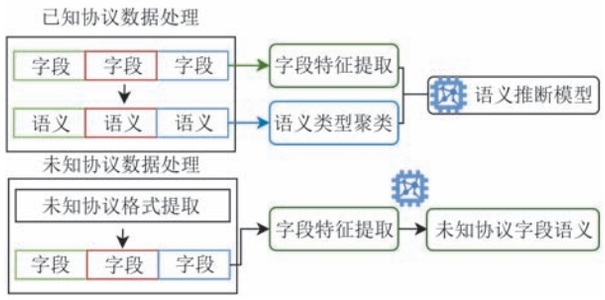

图4　协议语义推断模块

提取每个字段相应的特征，通过和已知语义类别之间的相似程度预测未知协议语义。

3.2.2　字段特征提取

字段特征提取在协议消息分组上使用循环神经网络从多个维度提取字段信息，并对提取信息加以综合来形成字段特征。字段信息包含字段本身和字段上下文，其中字段本身即字段所包含的连续字节，而字段上下文包含同一字段上下文和不同字段上下文，它们分别表示了两个不同维度的字段信息。图5(a)展示了字段特征提取方法的整体流程，(b)至(d)展示了各子流程的工作机制，下面将按照字段特征提取流程对细节进行介绍。

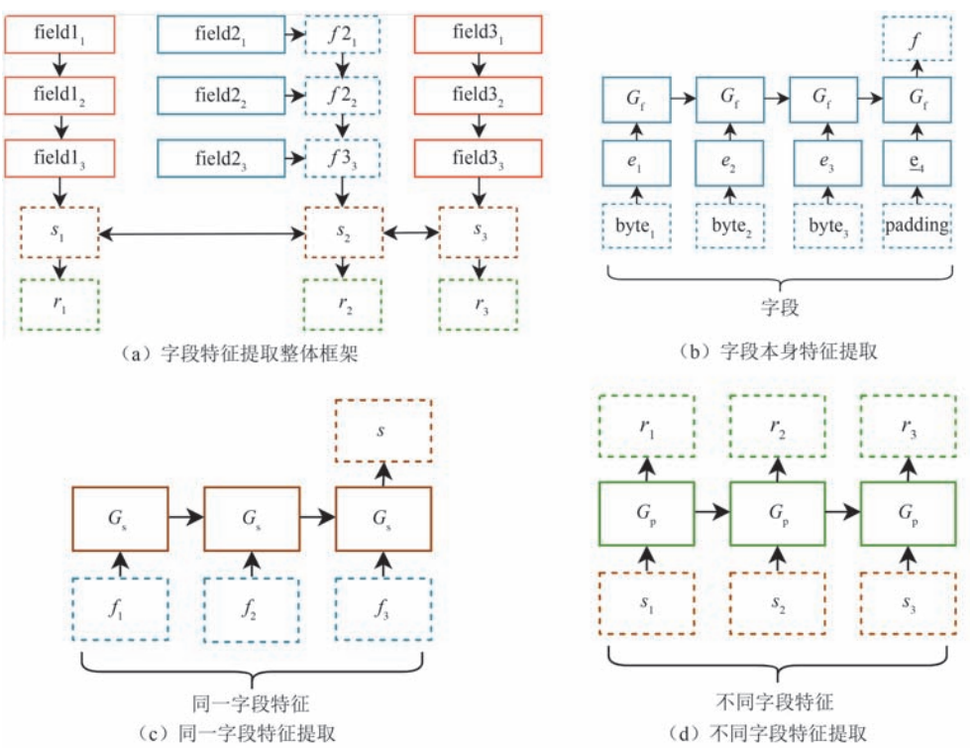

图5　字段特征提取方法及其子过程(子图(a)为字段提取方法的整体框架，子图(b)，(c)和(d)为字段特征提取方法的子过程)

(1)　字段内特征提取

字段特征提取首先提取字段本身的特征，字段本身特征提取过程如图5(b)所示，G_f表示GRU单元，G_f读取字段的连续字节序列并获得字段的特征

输出。由于字节序列是0到255之间离散值集合，为了更好地提取语义信息，需要先用权重矩阵将字节值映射到高维中的连续向量，获取字节值的嵌入向量e，权重矩阵的参数在训练中会得到更新。另外为了计算消息组的总体损失，要让消息组所有字段值同时输入到网络模型中，但字段的长度不同，需要对字段的连续字节值进行处理。为了解决这个问题，本文使用256作为字段填充，最终字段长度被填充为所有字段长度的最大值。这里注意，在训练过程中，填充值并不参与参数更新的计算，因此填充值并不影响字段本身特征的提取。

（2） 同一字段特征提取

在完成了所有字段本身的特征值提取后，需要对同一字段的上下文特征进一步进行提取。如图5(a)所示，以field2为例，不同消息中的field2字段分别提取为特征向量$f2_1, f2_2, f2_3$，然后将使用图5(c)中G_s单元将特征向量序列进一步做特征提取，得到同一字段聚合特征。其中G_s表示同一字段上下文特征提取所使用的GRU单元，它将消息分组中同一字段的特征提取向量作为输入，最终获得的隐藏层输出s作为同一字段上下文的特征提取值。

（3） 不同字段特征提取

如图5(a)所示，分别对field1、field2和field3完成同一字段特征提取后，会得到特征序列s_1, s_2, s_3，然后将使用图5(d)中的G_P单元进一步对特征序列进行处理，综合不同字段的上下文特征，得到最终聚合的字段提取特征r_1, r_2, r_3。这里注意，对于不同字段的上下文特征提取，并不是获得整个序列的信息，而是获取每一个字段考虑前后字段特征后的输出，因此这里G_P为双向的GRU单元，以便向前和向后双向获取特征，每一个G_P单元的输出作为对应字段的最终特征。

综上所述，可将G_f, G_s, G_P聚合为G_F，G_F代表了整个字段特征提取的过程，它以一个消息组的所有字段作为输入，最终得到特征序列$R=\{r_1, r_2, \cdots, r_m\}$，其中$m$代表消息中字段的个数，$r_i$代表每个字段相对应的提取特征。

3.2.3 字段语义聚合

对于已知协议，协议解析器会提供每个字段的语义描述，但是这种语义描述是细粒度的，无法很好地应用在未知协议语义推断中，因此需要将已知协议的细粒度语义聚合为语义类别。字段语义可以通过Wireshark[32]工具进行提取，但字段语义的描述是自然语言，需要使用自然语言处理(NLP)模型将字段语义提取为特征以便于进一步聚类。这里本文选择使用在许多大型语料库上预训练的all-MiniLM-L12-v2语言模型[33]，它能够将语义描述映射到384维的稠密向量空间中，将语义描述转化为句子嵌入。然后计算句子嵌入向量之间的余弦相似度就可以得到相似矩阵，再利用相似矩阵对字段语义进行聚类。

聚类的一个关键问题是如何确定分类数量，这个值可以被人工预先定义，但是这样需要人的经验判断，可能会存在误差，因此本文设计了一种能够自动找到分类数量的方法。具体来说，首先设置一个分类数量k的范围，然后使用范围内不同的k值基于相似矩阵进行k-means聚类[34]。k-means聚类会将n个输入分为k组($n>k$)，并最小化组内的平方距离。那么在执行k-means聚类后，就可以得到不同k值所对应组内距离的平方和c。一般来说，k值越大分类数量越多，更多的类别显然会造成更少的组内元素，那么c的值也会倾向于更小。但是k的值也不能过大，过大的k值会导致聚合语义类别的效果变差，因此需要找到合适的k值来平衡组内元素的相似性和聚类效果。

在计算机系统中，通常会达到一个点，这个点被称为“膝点”[35]，即增加某些可调参数的相对成本不再值得相应的性能收益。在这个问题中也具有相似的性质，因此可以在聚类数量k和组内平方距离和c构成的曲线上计算这个点作为最终分类数量。这里将曲线定义为f，则f在点x处的曲率为

$$K_f(x)=\frac{f''(x)}{\left(1+f'(x)^2\right)^{1.5}} \tag{4}$$

其中，f'和f''分别表示曲线的一阶和二阶导数，膝点的值为曲率最大位置所对应的x的值。图6直观展示了组内平方距离和c随k值变化的曲线(C为均值

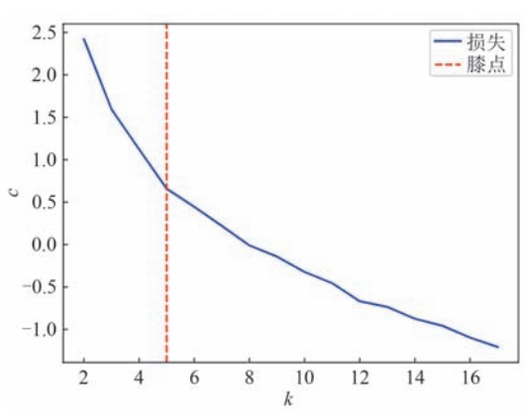

图6 组内距离平方和c随k变化曲线

归一化后的数值），曲线上k=5的位置被标记为膝点，它的值作为最终分类的数量。

3.2.4　模型训练及语义相似度推断

目前已经完成了字段特征提取和语义类型的聚类，得到了已知协议字段和语义类别之间的对应关系。那么在模型训练时，如果协议字段属于同一语义类别，它们的特征向量之间就应该具有更高的相似度。为了更好地对特征向量进行相似性表征，本文采用了对比学习的方法，其通过数据点比对来学习表征，在多种分类任务[36-37]中表现出很好的效果。具体来说，在每一轮的训练中首先随机选择一个字段提取特征作为锚点，同一类别的字段特征视为正样本对，而不同类别的字段特征视为负样本对，训练的目标就是让正样本对更加靠近锚点，负样本对远离锚点。因此损失函数应该对违背目标的数据点产生更大的损失，具体来说，给出字段的一组特征表示$R=\{\boldsymbol{r}_1,\boldsymbol{r}_2,\cdots,\boldsymbol{r}_n\}$，其中$r_i$被选择为锚点。$P$代表所有和$r_i$同类的正样本集合，$|P|$代表同类字段的个数，$N$代表所有的负样本集合，可以得到损失函数的公式：

$$L_t=\sum_{r_i\in R}\frac{-1}{|P|}\sum_{r_p\in P}\log\frac{\exp(\boldsymbol{r}_i\cdot\boldsymbol{r}_p/\tau)}{\sum_{r_n\in N}\exp(\boldsymbol{r}_i\cdot\boldsymbol{r}_n/\tau)}\tag{5}$$

其中，·表示向量之间的内积，τ表示温度系数。这里对于已经归一化的r_i来说，内积可以表示向量之间的余弦相似度。考虑到要将已知协议上的训练结果应用到未知协议中，但是未知协议和已知协议字段之间的差别可能很大，因此仅仅使用上面的损失函数有可能出现过拟合现象，造成模型的泛化能力不足，所以方案在训练中加入了数据增强的方法。

具体来说，每次训练迭代中，在训练数据集中选择一种协议构建其子消息数据作为增强数据。假设训练数据集中包含两类协议，在第一类协议的数据集中随机选择部分消息来构建子消息数据，那么对子消息数据集进行训练，就会得到一组特征集$R^*=\{r_1,r_2,\cdots,r_{n_1}\}$，其中$n_1$表示第一类协议中字段的个数。若$r_i\in R^*$被选为锚点，则将$R^*$中同类字段特征视为正向样本，标记为$r_i^+$。这里注意，对于非同类字段特征，除了$R^*$中的非同类字段特征外，还包含子消息数据集外所有的字段提取特征，此部分负样本数据集记为N^*，那么增强数据集上的损失函数就可以表示为：

$$L_s=-\sum_{r_i\in R^*}\log\frac{\exp(r_i\cdot r_i^+/\tau)}{\sum_{r_n\in N^*}\exp(r_i\cdot r_n/\tau)}\tag{6}$$

最终损失函数如公式(7)所示，结合了训练数据集和增强数据集两部分，其中λ为比例系数，通过这个公式就可以让同一类别的特征具有更高的相似性，同时也具有适度的泛化能力。

$$L=L_t+\lambda L_s\tag{7}$$

接下来利用损失函数对模型进行训练，经过训练后，模型就可以按照语义类别对不同字段的特征提取向量进行聚类，同一类别的字段其特征提取向量将具有更高的相似性。然后对同一类别字段的特征提取向量进行归一化并计算中心点（向量平均值）作为语义类别的参照。

在对未知协议语义进行相似度语义推断时，首先对未知协议进行数据处理，应用格式提取工具得到未知协议字段并将协议消息划分为消息分组。之后利用训练模型在消息分组上提取未知协议字段特征，计算得到字段特征向量和语义类别中心点之间的相似度，根据相似度分数便可以推断未知协议字段语义。

4　实验评估

本文的未知协议逆向分析方法主要包含协议格式提取和协议语义推断两个模块，本节针对这两个模块分别设计了实验进行了测试。

4.1　未知协议格式提取

4.1.1　实验设置

协议格式提取方法需要通过已知协议进行训练。为了充分学习协议特征需要大量数据，然而，数据量的增加会显著延长训练时间，且部分网络协议的公开可用数据量有限。因此，本文基于互联网公开流量库[38-40]筛选协议样本，并系统尝试了多种数据分布策略，包括均衡采样、过采样低频协议及非均衡原始分布等。实验验证表明，采用非均衡原始分布时，能够较好地平衡训练时间和格式推断效果，故最终选取此分布训练。训练数据具体情况如表1所示：

表1　训练数据情况

协议	数量/条	协议	数量/条
ARP	12 480	ICMP	9856
BGP	200	ICMPV6	3840
DNS	12 672	IP	31 420
ESP	14 700	TCP	31 488
UDP	30 976	OSPF	576
NBDGM	2624	SMB	3520

将训练数据按照8:2划分训练集和验证集，使用RMSprop梯度下降算法，初始学习率设置为0.0001，*momentum*=0.8图7展示了训练数据和验证数据损失率随训练回合变化情况，可以看出，经过50个回合的训练，训练数据和验证数据的损失都已经下降到0.35左右，说明模型在训练数据集和验证数据集上均表现良好。

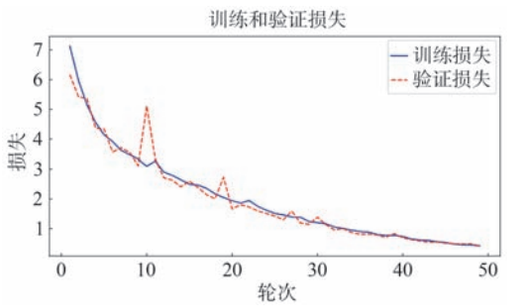

图7 损失率变化情况

4.1.2 格式提取效果分析和比较

为了对未知协议的字段划分结果更好地进行评价，本文定义了精确度，覆盖率和F1值三个评价指标。

首先需要定义计算这三个评价指标的变量，图8是这三个变量所对应位置的直观展示：匹配位置(TP)，即正确匹配真实情况的位置个数；不匹配位置(FP)，预测结果不正确的位置个数；丢失位置(FN)，预测未知丢失的个数。

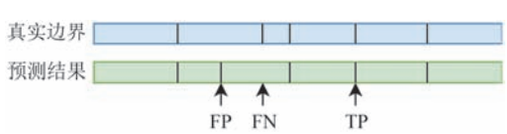

图8 真实值和预测结果对应关系

根据变量便可以计算得到下面的评价指标：

• 精确度(*Prec.*)：匹配位置的数量占所有预测位置的数量 $Prec=\frac{TP}{TP+FP}$

• 覆盖率(Cov.)：匹配位置的数量占所有真实分割位置的数量 $Cov=\frac{TP}{TP+FN}$

• F1值：同时考虑和覆盖率的结果表征值 $F1=\frac{2\times Prec\times Cov}{Prec+Cov}$

本文选择了Modbus、S7comm、DNP3和NBNS四种协议来测试协议格式提取模块的效果。这四种协议可以通过tshark工具进行解析以便获取真实情况进行对照，但在测试中只将相关的数据流量提供给本文的格式提取方案，模型获取不到其它任何先验信息，因此对于模型来说，协议的规范是完全未知的。

本文在4种协议上分别对BPREN的协议格式提取效果进行测试，并计算相应的评价指标。为了更好地展示本方案的效果，这里选用Netzob[17]和NEMESYS[41]两种开源的格式提取工具作为对比。Netzob和NEMESYS分别是基于对齐和基于概率的格式提取方法，选择它们作为代表能够很好地展现不同方案之间的优劣性。

为了保证对比一致性，本文使用相同的数据对不同方案进行了测试。在上述四种协议上的测试结果如表2所示。这里注意，在Netzob和NEMESYS中均存在需要预先设定的参数，如Netzob中的minEquivalence和NEMESYS中的sigma，本文在测试中对这些参数进行迭代，并选择了最好的预测效果作为对比。从表中可以看出，本文方案在四种协议上各项指标均优于NEMESYS方案。同时和Netzob方案相比，虽然在DNP3和NBNS协议上预测准确率略低，但是对于真实预测情况的覆盖率更高，因此在综合评价指标F1上实现反超。综合四种协议的评估结果，本文方案相较于Netzob，NEMESYS方法在综合评价指标F1上平均分别有19%和49%的提升。

为了更直观地展示本文方法在协议格式提取上的有效性，本文选取了一个典型的协议消息作为实例，并与Netzob和NEMESYS的推断结果进行了对比。图9展示了三种方法对S7comm协议头部字段的推断结果，空格表示字段边界真实值，实线和虚线分别表示未识别的边界和错误识别的边界。

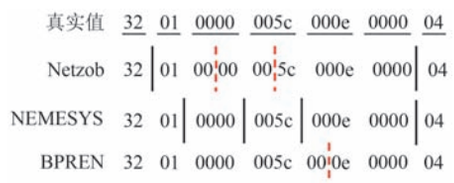

图9 S7comm协议头部字段格式提取效果对比

从实验结果中可以发现，基于序列对齐的Netzob方法对于数据样本的依赖性较高，难以有效推断上下文变化不明显的字段，并且对于字段内部部分字节变化的字段，容易出现误报的情况。NEMESYS方法虽然在字段边界推断上表现出更

高的准确性，但其对真实字段的覆盖率相对较低。这可能是因为它依赖于对特定统计特征的分析，而无法从整体上捕捉到协议消息中的特征模式。相比之下，本文模型基于深度学习，能够从大量协议数据中自动学习并提取字段特征，在面对缺乏内部信息的未知协议时，依然能够将学习到的相似字段特征进行迁移，因此综合表现更优。

表2　不同方案格式提取效果对比

协议	BPREN			Netzob			NEMESYS		
	Prec.	Cov.	F1	Prec.	Cov.	F1	Prec.	Cov.	F1
Modbus	**1.00**	**0.60**	**0.75**	0.75	0.50	0.60	**1.00**	0.5	0.67
S7comm	**0.96**	**0.82**	**0.88**	0.83	0.63	0.71	0.83	0.31	0.45
DNP3	0.55	**0.50**	**0.53**	**0.57**	0.40	0.47	0.50	0.30	0.37
NBNS	0.61	**1.00**	**0.76**	**0.80**	0.57	0.67	0.60	0.43	0.50

本文还对不同方案之间的格式提取效率进行了对比，分别选择了100，200，500和1000条协议消息进行了测试。对于本文方案，在效率测试中不用考虑离线训练所耗费的时间，但参数和模型加载时间被计算在内。从表3的测试结果可以看出，本文方案和基于概率的NEMESYS方法效率总体相差不大，虽然Netzob方法在对少量消息做格式提取时相对本文方案效率略高，但由于Netzob的序列对齐算法的递归特性，其在协议数量增加时效率迅速下降，远低于NEMESYS和本文方案。

表3　不同方案效率对比

协议数量	方案运行时间(s)		
	BPREN	Netzob	NEMESYS
100	1.14	1.06	0.96
200	1.18	4.05	1.84
500	3.13	84.07	3.01
1000	7.25	539.27	7.13

4.2　未知协议语义推断

4.2.1　实验设置

对于协议语义推断方案，本文使用了7种不同的二进制协议进行测试，其中包含3种网络协议（DHCP，SMB和SMB2）和四种工业控制协议（Modbus，DNP3，S7comm和OMRON）。每个协议使用约200条消息作为输入生成随机消息分组，每个消息分组的大小被设置为10。

4.2.2　未知协议语义提取效果分析

目前的语义推断方案大多集中于未知协议的部分字段，最后推断的结果也往往是唯一对应，而本文的语义推断模块可以基于相似性对所有字段的语义进行推断，同时语义推断结果依赖于协议格式提取模块的字段划分，因此不同的方案之间不容易直接比较。

由于本文的方案是基于相似性的推断，这里本文基于相似性设置了两个评价指标对语义推断的准确性进行评估：

Acc1：只考虑预测输出相似度最高的语义类别的准确性，若相似度最高的预测语义类别和真实值一致，则视为预测准确。

Acc2：考虑预测输出相似度最高和次高语义类别的准确性，具体来说，若相似度最高和次高的语义类别其中一项和真实值一致，就视为预测准确。

本文在本部分测试评估中重点关注协议语义推断模块的预测结果，因此直接使用了真实字段作为了测试输入。而在实际的未知协议逆向分析过程中，一般需要协议格式提取的结果作为字段的输入值，不同格式提取方法的结果有可能影响最终语义推断的准确度。

本文方法的语义推断依赖于训练样本和测试数据之间的语义相似度，因此本文在测试中将样本按照不同的协议分组进行实验评估。首先在网络协议组上进行测试，对其中一种协议测试时，使用协议组内的其它协议进行训练。从表4中可以看到，在网络协议组上训练，Acc2准确率均在70%左右，说明结合最高和次高相似度的语义类别能够对未知协议字段语义做出很好的预测。

表4　网络协议组测试结果

测试协议	K	Acc1	Acc2
DHCP	4	61.54%	69.23%
SMB	5	58.97%	76.92%
SMB2	4	41.03%	66.67%

另外本文在工业控制协议组上也使用了相同的规则进行了测试，如对Modbus协议进行测试时，使用OMRON，S7comm和DNP3协议进行训练，最终的结果如表5所示。可以看到在工控协议Acc1和

表5　工业控制协议组测试结果

测试协议	K	Acc1	Acc2
Modbus	5	71.43%	83.67%
S7comm	4	49.28%	65.79%
DNP3	5	55.10%	69.39%
OMRON	4	65.53%	66.29%

Acc2的平均值分别达到了60%和71%，说明通过最高相似度推断类别就可以较好地获取未知协议字段的语义。

为了更好地说明本文语义推断的结果，本文选择了S7comm协议进行实例分析，表6直观展示语义推断的过程以及语义推断的效果。

表6 S7comm协议字段语义推断结果

字段	真实语义	最高相似度推断结果
1	Protocol id	‘Unit Identifier’,‘Service ID’,‘Protocol Identifier’,‘Transaction Identifier’,‘Register Number’
2	ROSCTR	‘Response code’,‘Application Control’,‘Transport Control’,‘Control’,‘Return code’
3	Redundancy Identification (Reserved)	‘Response code’,‘Application Control’,‘Transport Control’,‘Control’,‘Return code’
4	Protocol Data Unit Reference	‘Byte Count’,‘Length’,‘No. of total words’,‘Item count’
5	Parameter length	‘Byte Count’, ‘Length’,‘No. of total words’,‘Item count’
6	Data length	‘Reserved’,‘Kind of DM’,‘Object’,‘OMRON ICF Field’

表6中包含了S7comm协议头部字段的最高语义相似度推断结果，其中字段1，2和5的推断结果与真实语义高度匹配。例如，ROSCTR字段是S7comm协议中的请求/响应操作码控制字段，用于区分报文类型。模型将其推断为“Control”、“Return code”等已知字段语义，这些结果反映了该字段的控制和状态反馈功能，为研究人员推断其具体功能提供了依据。对于Protocol Data Unit Reference字段，虽然其最高语义相似度推断结果与真实语义不匹配，但通过次高的语义相似度类别中的语义，如“Unit Identifier”和“Transaction Identifier”，研究人员仍可结合其经验推断出该字段的功能。从结果中可以看出，本文提出的方法能够有效识别字段语义，并为协议分析提供有价值的线索。

此外，实验中发现训练与测试协议在语义上的相似性对模型推断性能存在影响。当使用语义类别上相近的协议进行训练时，推断结果的准确率较高；反之，当测试协议与训练协议语义类别存在差异时（如使用网络协议训练推断工业控制协议），推理准确度会明显下降。这一现象的主要原因在于不同协议语义之间的差异性，尽管深度学习模型能够从已知协议中学习和迁移字段特征，但协议的语义与具体的协议功能和应用场景高度相关，模型很难推断出不同协议领域中特有的字段语义。因此，为了提升模型的推断效果，在实际应用中应尽量使用语义相似或同一协议簇的样本进行训练。

5 总结和展望

未知网络协议逆向分析是关乎网络安全的重要挑战，分析人员需要对网络协议的格式、语义和状态机等进行恢复以帮助安全分析。但目前未知网络协议逆向分析存在很多难点，首先由于二进制协议不具备明显的分隔符且缺乏连续可读字节，和文本协议相比更难提取其协议规范。现有的二进制协议逆向方法在提取协议格式时往往需要人工设置参数，同时很多方案只针对部分特殊协议，缺乏通用性。此外对于二进制协议的语义推断方案一般只能推断未知协议中的部分特殊字段，对复杂的未知字段不能很好地解决，在自动化程度方面同样也存在欠缺。

针对上述存在的问题，本文重点结合深度学习手段实现了未知协议格式方法BPREN，分别设计了协议格式提取和语义推断模块。在协议格式提取模块中，本文利用U-Net图像语义分割网络对已知协议消息图像特征进行学习，利用训练模型提取未知协议的协议格式，并对多图像预测结果进行合理聚合，提升了格式提取结果的准确性。在协议语义推断模块中，本文使用GRU循环神经网络对已知协议字段特征进行了多维度的提取，并结合抽象语义聚类对模型进行训练，最后基于字段提取特征的相似性对未知协议字段语义进行推断，同时在训练过程中加入了数据增强方案，使模型能够更好对未知协议字段语义做出预测。

论文在实验测试部分对BPREN方法进行了评估，对于协议格式提取模块，本文分别在准确度，覆盖率和F1值（综合评价指标）上对本文方案进行了测试，并和Netzob，NEMESYS两种代表性方案进行了对比。在Modbus，S7comm，DNP3和NBNS 4种协议上，本文方案预测结果对真实边界的覆盖率均优于Netzob和NEMESYS，在F1值上平均分别有19%和49%的提升。另外本文方案在处理较大数量的协议流量数据时效率也表现出色，远好于Netzob并和NEMESYS基本一致。

对于协议的语义推断模块，本文选取了7种不

同的工控和网络二进制协议进行测试。首先验证了本文方案对协议全字段语义的推断效果，之后针对所有样本数据和同类协议分别进行了测试。使用同类网络协议进行测试时，在网络协议组上Acc1和Acc2准确率平均分别为54%和71%，在工业控制协议组上Acc1和Acc2准确率平均分别为60%和71%。通过对比实验结果可以看出，当使用相似度较高的协议进行训练时，本文语义推断方案能够对未知协议的字段语义做出较为有效的推测。

基于本文的未知协议逆向分析方法，研究人员可以分别完成未知二进制协议的格式提取和语义推断工作，方案能较为高效地恢复未知协议的部分协议规范，同时自动化程度较高，通用性也比强，具有应用价值。

本文方法也存在一些局限性。如在格式提取模型的上采样过程中，本文有目的地对纵向特征进行了压缩，这提升了模型格式提取的速度，但也可能导致部分纵向关联性信息的损失。其次，本文对长消息采用截断处理，尽管通过递归机制确保了数据都能被分析，但若截断恰好发生在某个字段的内部，可能会导致该字段的特征被分割，从而在一定程度上影响格式提取的准确性。在今后的工作中，可以对此模型进行改进，比如加入序列建模方法来捕获消息序列间的纵向特征，并将融合后的特征输入解码器，以期在保持高效性的同时，进一步提升对复杂协议的分析能力。

参 考 文 献

[1] China Internet Network Information Center(CNNIC). The 56th Statistical Report on China's Internet Development. Beijing, 2025 (in Chinese)
(中国互联网络信息中心.第56次中国互联网络发展状况统计报告.北京，2025年7月21日)

[2] Hoque E, Chowdhury O, Chau S Y, et al. Analyzing operational behavior of stateful protocol implementations for detectingsemantic bugs//Proceedings of the 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Denver, USA, 2017: 627-638

[3] Borgolte K, Kruegel C, Vigna G. Delta: automatic identification of unknown web-based infection campaigns//Proceedings of the 2013 ACM SIGSAC conference on Computer & Communications Security. Berlin, Germany, 2013: 109-120

[4] Österlund S, Razavi K, Bos H, et al. Parmesan: sanitizer-guided greybox fuzzing//Proceedings of the 29th USENIX Security Symposium. Virtual, USA, 2020: 2289-2306

[5] Jain V, Rawat S, Giuffrida C, et al. Tiff: using input type inference to improve fuzzing//Proceedings of the 34th Annual Computer Security Applications Conference. San Juan, USA (Puerto Rico), 2018: 505-517

[6] Chen Y, Mu D, Xu J, et al. Ptrix: efficient hardware-assisted fuzzing for cots binary//Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security. Auckland, New Zealand, 2019: 633-645

[7] Cao Y, Shoshitaishvili Y, Borgolte K, et al. Protecting web-based single sign-on protocols against relying party impersonation attacks through a dedicated bi-directional authenticated secure channel//Proceedings of the Research in Attacks, Intrusions and Defenses: 17th International Symposium, RAID 2014. Gothenburg, Sweden, 2014: 276-298

[8] Starnberger G, Kruegel C, Kirda E. Overbot: a botnet protocol based on kademlia//Proceedings of the 4th international conference on Security and privacy in communication netowrks. Istanbul, Turkey, 2008: 1-9

[9] Antonakakis M, April T, Bailey M, et al. Understanding the mirai botnet//Proceedings of the 26th USENIX Security Symposium. Vancouver, Canada, 2017: 1093-1110

[10] Jiang J, Zhang X, Wan C, et al. Binpre: enhancing field inference in binary analysis based protocol reverse engineering//Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security. Salt Lake City, USA, 2024: 3689-3703

[11] Ma R, Zheng H, Wang J, et al. Automatic protocol reverse engineering for industrial control systems with dynamic taint analysis. Frontiers of Information Technology & Electronic Engineering, 2022, 23(3): 351-360

[12] Sun Y, Li Z, Lv S, et al. Spenny: Extensive ICS protocol reverse analysis via field guided symbolic execution. IEEE Transactions on Dependable and Secure Computing, 2022, 20(6): 4502-4518

[13] Qu Y, Fang D, Wang Z, et al. ICEPRE: ICS protocol reverse engineering via data-driven concolic execution. Proceedings of the ACM on Software Engineering, 2025, 2(ISSTA): 2384-2406

[14] Li Junchen, Cheng Guang, Yang Gangqin. A review of private protocol reverse engineering technology based on network traffic. Journal of Computer Research and Development, 2023, 60(01): 167-190 (in Chinese)
(李峻辰，程光，杨刚芹.基于网络流量的私有协议逆向技术综述.计算机研究与发展，2023，60(01):167-190)

[15] Beddoe M A. Network protocol analysis using bioinformatics algorithms. Toorcon, 2004, 26(6): 1095-1098

[16] Cui W, Kannan J, Wang H J. Discoverer: Automatic protocol reverse engineering from network traces//Proceedings of the USENIX Security Symposium(USENIX Security 07. Boston, USA, 2007: 1-14

[17] Bossert G, Guihéry F, Hiet G. Towards automated protocol reverse engineering using semantic information//Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security. Kyoto, Japan, 2014: 51-62

[18] Luo Z, Liang K, Zhao Y, et al. Dynpre: Protocol reverse

engineering via dynamic inference//Proceedings of the Network and Distributed System Security Symposium. San Diego, USA, 2024: 1-18

[19] Wang Y, Zhang Z, Yao D, et al. Inferring protocol state machine from network traces: a probabilistic approach//Proceedings of the 9th International Conference on Applied Cryptography and Network Security. New York, USA, 2011: 1-18

[20] Ye Y, Zhang Z, Wang F, et al. NetPlier: Probabilistic networkprotocol reverse engineering from message traces//Proceedings of the 28th Network and Distributed Systems Security Symposium. Virtual, USA, 2021:2631-2647

[21] Liang K, Luo Z, Zhao Y, et al. MDIplier: Protocol format recovery via hierarchical inference//Proceedings of the 2024 IEEE 35th International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2024: 547-557

[22] Zhao R, Liu Z. Analysis of private industrial control protocol format based on lstm-fcn model//Proceedings of the 2020 International Conference on Aviation Safety and Information Technology. Weihai, China, 2020: 330-335

[23] Kiechle V, Börsig M, Nitzsche S, et al. PREUNN: Protocol reverse engineering using neural networks//Proceedings of the 9th International Conference on Information Systems Security and Privacy(ICISSP).Virtual, USA,2022: 345-356

[24] Zhang W, Meng X, Zhang Y. Dual-track protocol reverse analysis based on share learning//Proceedings of the IEEE INFOCOM 2022-IEEE Conference on Computer Communications. Virtual, USA, 2022: 51-60

[25] Cui W, Paxson V, Weaver N, et al. Protocol-independent adaptive replay of application dialog//Proceedings of the 13th Network and Distributed System Security Symposium. San Diego, USA, 2006: 1-15

[26] Bermudez I, Tongaonkar A, Iliofotou M, et al. Automatic protocol field inference for deeper protocol understanding//Proceedings of the 2015 IFIP Networking Conference. Toulouse, France, 2015: 1-9

[27] Chandler J, Wick A, Fisher K. Binaryinferno: A semantic-driven approach to field inference for binary message formats//Proceedings of the 30th Network and Distributed System Security Symposium. San Diego, USA, 2023: 1-18

[28] Kleber S, Kargl F. Poster: Network message field type recognition//Proceedings of the 2019 ACM SIGSAC Conf on Computer and Communications Security. New York, USA, 2019: 2581-2583

[29] Wang Qun, Sun Zhonghua, Wang Zhangquan, et al. A practical format and semantic reverse analysis approach for industrial control protocols. Security and Communication Networks, 2021, 2021: 6690988

[30] Yang D, Yao Y, Shan Y, et al. Patty: Pattern series-based semantics analysis for agnostic industrial control protocols. IEEE Transactions on Information Forensics and Security, 2025,20: 5478-5491

[31] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany, 2015: 234-241

[32] Wireshark, https://www.wireshark.org/, (2022-03-06), 2024-04-15

[33] modelSentence-transformers, https://huggingface.co/sentence-transformers/all-MiniLM-L12-v2, (2024-3-26), 2024-04-13

[34] Saro, Kavita. Review: study on simple k mean and modified K mean cluste ring technique. International Journal of Computer Science Engineering and Technology,2016,6(7):279-281

[35] Satopaa V, Albrecht J, Irwin D, et al. Finding a "kneedle" in a haystack: detecting knee points in system behavior//Proceedings of the 2011 31st International Conference on Distributed Computing Systems Workshops. Minneapolis, USA, 2011: 166-171

[36] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations//Proceedings of the International conference on machine learning. PMLR, 2020: 1597-1607

[37] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning. Adv ances in neural information processing systems, 2020, 33: 18661-18673

[38] TablePcap, https://tshark. dev/search/pcaptable, (2019-09-09), 2024-04-13

[39] Public PCAP files for download, https://www. netresec. com, (2024-03-18), 2024-04-13

[40] LBNL/ICSI Enterprise Tracing Project, http://www.icir.org/enterprise-tracing/download.html, (2013-7-30), 2024-04-13

[41] Kleber S, Kopp H, Kargl F. {NEMESYS}: Network message syntax reverse engineering by analysis of the intrinsic structure of individual messages//Proceedings of the 12th USENIX Workshop on Offensive Technologies (WOOT 18). Baltimore, USA, 2018: 1-13

LI Yu-Qing, Ph. D., associate professor, Ph. D. supervisor. Her research interests include network security, distributed system security.

LI Zhuo-Qun, M. S. candidate. His research interests include network security and IoT security.

CHEN Jing, Ph. D., professor. Ph. D. supervisor. His research interests include network security, distributed system security.

HE Kun, Ph. D. associate professor. Ph. D., supervisor. His research interests include applied cryptography, network security.

DU Rui-Ying, Ph. D., professor, Ph. D. supervisor. Her research area is network security, privacy protection.

SUN Xi-Ping, Ph. D. candidate. Her research area is mobile security.

WU Cong, Ph. D., Post-doctoral Researcher. His main research interests include distributed system security.

Background

This research focuses on the problem of reverse engineering unknown binary protocols within the field of network and information security. Protocol reverse engineering (PRE) aims to reconstruct the format, semantics, and behavior of undocumented or proprietary communication protocols. With the rapid expansion of the Internet of Things (IoT), an increasing number of private protocols have emerged, especially in industrial control systems, smart devices, and autonomous systems. These protocols are often undocumented and lack publicly available specifications, posing significant challenges for security analysis, intrusion detection, fuzz testing, and malware behavior analysis.

Currently, two mainstream approaches dominate the international research landscape: binary-based and network-trace-based PRE. Binary-based methods analyze executable files through symbolic execution, taint tracking, or static analysis, but they require access to firmware or binary code, which is often unavailable or obfuscated in practice. In contrast, network-trace-based methods analyze captured traffic to infer protocol structures using heuristic, statistical, or machine learning techniques. However, many existing methods suffer from limited automation, poor generalization to unseen protocols, and heavy dependence on manual assumptions or high-quality traffic samples.

This paper proposes BPREN, a deep learning-based framework that addresses the limitations of prior approaches by leveraging known protocol traffic to learn transferable representations. It introduces a dual-module design consisting of a U-Net-based protocol format extraction module and a GRU-based semantic inference module. The system achieves significant performance improvements in terms of accuracy, generalization, and automation across a variety of real-world binary protocols.

The research is supported by the following major national and provincial projects:

(1) National Natural Science Foundation of China (NSFC) (Nos. 62302343,62472323, 62172303)

(2) Key R&D Projects of Hubei Province (No. 2024BAB018)

(3) Wuhan Scientific and Technical Achievements Project (No. 2024030803010172)

(4) Key R&D Program of Shandong Province (No. 2022CXPT055)

(5) Wuhan City Joint Innovation Laboratory for Next-Generation Wireless Communication Industry Featuring Satellite-Terrestrial Integration(No. 4050902040448)

The research group has long been engaged in protocol analysis, IoT security, and traffic behavior modeling. Previous works include binary code analysis and semantic extraction in ICS environments. This study constitutes a critical subtask of large-scale intelligent network security infrastructure projects and directly contributes to the automatic modeling and understanding of unknown communication protocols.