计算机科学
COMPUTER SCIENCE

基于生成对抗网络与变异策略结合的网络协议漏洞挖掘方法

庄园，曹文芳，孙国凯，孙建国，申林山，尤扬，王晓鹏，张云海

引用本文

庄园，曹文芳，孙国凯，孙建国，申林山，尤扬，王晓鹏，张云海. 基于生成对抗网络与变异策略结合的网络协议漏洞挖掘方法[J]. 计算机科学，2023，50(9)：44-51.

ZHUANG Yuan，CAO Wenfang，SUN Guokai，SUN Jianguo，SHEN Linshan，YOU Yang，WANG Xiaopeng，ZHANG Yunhai. Network Protocol Vulnerability Mining Method Based on the Combination of Generative AdversarialNetwork and Mutation Strategy [J]. Computer Science，2023，50(9)：44-51.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230600013

基于生成对抗网络与变异策略结合的网络协议漏洞挖掘方法

庄　园[1]　曹文芳[1]　孙国凯[1]　孙建国[2]　申林山[1]　尤　扬[3]　王晓鹏[3]　张云海[3]
1 哈尔滨工程大学计算机科学与技术学院　哈尔滨 150001
2 西安电子科技大学杭州研究院　杭州 311231
3 绿盟科技集团股份有限公司　北京 100089
(zhuangyuan@hrbeu.edu.cn)

摘　要　随着信息化和工业化的深度融合,工业物联网网络协议安全问题日益突出。现有网络协议漏洞挖掘技术以特征变异和模糊测试为主,存在依赖专家经验和无法突破未知协议的局限。针对工业物联网协议的漏洞挖掘挑战,文中从漏洞检测规则的自动化分析与生成展开研究,提出基于生成对抗网络与变异策略结合的网络协议漏洞挖掘方法。首先,采用一种基于生成对抗网络的网络协议分析模型,通过对报文序列进行深层信息挖掘,提取报文格式及相关特征,实现对网络协议结构的识别。然后,结合基于变异算子库指导的迭代变异策略,构建有导向性的测试用例生成规则,缩短漏洞发现的时间;最终,形成面向未知工控网络协议的自动化漏洞挖掘方法,满足现有工控应用领域对协议自动化漏洞挖掘的需求。基于上述方法,对两种工控协议(Modbus-TCP和S7)进行测试,并对生成用例的测试接收率、漏洞检测能力、用例生成时间及其多样性方面进行了评估。实验结果表明,所提方法在TA指标上高达89.4%,本方法检测模拟系统ModbusSlave的AD指标为6.87%,缩短了有效用例的生成时间,提升了工控协议漏洞挖掘的效率。

关键词:生成对抗网络;变异策略;模糊测试;漏洞挖掘;网络协议
中图法分类号　TP393

Network Protocol Vulnerability Mining Method Based on the Combination of Generative Adversarial Network and Mutation Strategy

ZHUANG Yuan[1],CAO Wenfang[1],SUN Guokai[1],SUN Jianguo[2],SHEN Linshan[1],YOU Yang[3],WANG Xiaopeng[3] and ZHANG Yunhai[3]
1 College of Computer Science and Technology,Harbin Engineering University,Harbin 150001,China
2 Hangzhou Institute of Technology,Xidian University,Hangzhou 311231,China
3 NSFOCUS Technologies Group Co.,Ltd.,Beijing 100089,China

Abstract　With the deep integration of informatization and industrialization,the security issues of industrial Internet of things (IIoT) network protocols are becoming increasingly prominent. Existing network protocol vulnerability mining techniques mainly rely on feature variation and fuzzy testing,which have the limitations of depending on expert experience and cannot overcome the challenges posed by unknown protocols. To address the vulnerability mining challenges in IIoT protocols,this paper conducts research on the automation analysis and generation of vulnerability detection rules and proposes a network protocol vulnerability mining method based on a combination of generative adversarial networks(GANs) and mutation strategies. Firstly,a network protocol analysis model based on GANs is employed to conduct deep information mining on message sequences,extract message formats,and related features,enabling the recognition of network protocol structures. Then,by combining a guided iterative mutation strategy with a mutation operator library,directed test case generation rules are constructed to reduce the time for vulnerability discovery. Ultimately,an automated vulnerability mining method for unknown industrial control network protocols is developed to meet the demand for protocol automated vulnerability mining in the existing industrial control application domain. Based on the above-mentioned approach,we conduct tests on two industrial control protocols(Modbus-TCP and S7) and evaluate them

到稿日期:2023-05-31　返修日期:2023-07-22
基金项目:CCF-绿盟科技"鲲鹏"基金(CCF-NSFOCUS 2021014);2022年工业互联网创新发展工程——工业互联网数据安全检测响应与溯源项目(TC220H055);中央高校基本科研业务费专项资金(3072022TS0604);西安电子科技大学杭州研究院概念验证基金项目(XJ2023230024)
This work was supported by the CCF-NSFOCUS(2021014),2022 Industrial Internet Innovation and Development Project — Industrial Internet Data Security Detection Response and Traceability Project (TC220H055), Fundamental Research Funds for the Central Universities (3072022TS0604) and Concept Foundation of Hangzhou Institute of Technology,Xidian University(XJ2023230024).
通信作者:申林山(shenlinshan@hrbeu.edu.cn)

in terms of test coverage, vulnerability detection capability, test case generation time, and diversity. Experimental results show that the proposed method achieves a remarkable 89.4% on the TA index. The AD index, which measures the ability to detect vulnerabilities in the simulated ModbusSlave system, reaches 6.87%. Additionally, the proposed method significantly reduces the time required for generating effective test cases, thereby enhancing the efficiency of industrial control protocol vulnerability discovery.

Keywords Generative adversarial network, Mutation strategy, Fuzzing test, Vulnerability mining, Network protocol

1　引言

工控协议指在工控环境下，双方实体完成通信或服务所需要遵守的规则及约定。近年来，随着工控产品和系统的普及，出现了许多网络安全事件，例如臭名昭著的震网病毒[1]、席卷全球的勒索病毒[2]等都给社会造成了重大损失。同时，网络上恶意攻击者利用工控网络漏洞对互联网上的设备发动远程攻击[3]，直接影响了整个互联网的安危。因此，工业物联网安全问题引起了世界各国的广泛关注。

漏洞挖掘技术指综合应用各种技术和工具，尽可能地发现软件程序、网络协议等存在安全漏洞的过程。据调查统计，传统工控网络协议漏洞挖掘方法主要采用的是逆向分析技术、渗透测试技术、模糊测试技术等。其中，模糊测试技术通常是根据需要测试的软件或协议规范来构建有效的测试用例，利用恶意输入对目标进行测试以引起崩溃或异常等行为，进而发现软件或者协议中存在的漏洞[4]。上述方法主要应用于当前已知的工控网络协议，但是由于存在投入成本高、执行耗时长、容易出错、针对性不足、迁移性差以及漏洞检测效率低等问题，这些方法难以实现智能、高效的漏洞挖掘。

因此，本文提出了一种基于生成对抗网络与变异策略结合的网络协议漏洞挖掘方法，旨在解决当前工控网络协议漏洞挖掘过程中遇到的问题。其中，利用生成对抗网络能够减少人工分析协议构造测试用例带来的不客观性问题，同时降低构造模糊测试用例的时间成本与人工成本，提高整个模糊测试过程的效率；依据工控协议报文字段特点以及现有漏洞规则的已知特征，设计指导性的变异策略与变异方法，在可控范围内高效、快速地指导生成有效用例，从而实现更精准、高效的漏洞挖掘。本文采用改进的多生成器生成对抗网络(DMGAN)产生测试用例，并且与线下快速变异策略结合的方式解决工控网络协议漏洞挖掘面临的难题，最后通过对Modbus-TCP协议与S7协议进行模拟仿真实验，验证了所提算法在模糊测试中测试用例数据多样性、测试系统接收率、故障的触发次数等性能指标上的有效性，以及对生成测试用例时间的改进。实验结果表明，利用DMGAN模型能够有效降低人工分析的时间和成本，在不影响漏洞检测效果的情况下，有效减少人工输入测试用例以及人工分析带来的不确定性，而变异策略能指导模型更加快速地找到有效的测试用例。综上，本文的贡献包括以下3个方面：

1)提出了基于DMGAN模型的网络协议漏洞挖掘方法，通过模型学习工控样本数据的报文格式，从而减少在工控系统测试过程中投入的人工成本和时间，提升测试效果。

2)根据线下分析工控协议报文字段的特点以及现有漏洞的触发原理，提出基于变异算子库和数据包规则的方法，该方法旨在设计有导向性的线下快速变异策略与方法来指导有效测试用例的生成，达到更加快速发现有效用例的效果。

3)采用迭代反馈的模式对测试用例进行不断的筛选和线下变异指导，结合模糊测试结果对变异策略与方法做进一步指导，同时将变异后的用例送入模型进行训练与测试，实现涉及层面更深、更广的测试，挖掘出更多的潜在漏洞。

2　相关工作

2.1　深度学习与生成对抗网络

近年来，机器学习和深度学习在技术上取得了重大突破。特别是深度学习的发展，使计算机具备了强大的感知能力；同时，深度学习也吸引了技术界的关注，并展现出巨大的应用前景。在游戏、机器人、机器翻译、语音识别、自动驾驶、导航、入侵检测、多智能体协作和推荐系统等领域中，深度学习已经实现了可与人类媲美甚至超越人类的表现。

在技术飞速发展的过程中，研究者开始将研究方向从机器感知转向机器创造，利用机器学习的生成技术，使机器具备创造新事物的能力。生成对抗网络的诞生，打破了人们对传统生成模型的理解，目前已经取得了令人称赞的成果。对抗网络的技术是人工智能领域新的里程碑。“生成对抗网络之父” Goodfellow 在蒙特利尔大学跟随深度学习顶级大师 Yoshua 深造，在生物学家 Leigh 的“红皇后假说”的启发下产生了对抗网络的构想，随后经历多次尝试后在对抗网络生成图像方面取得了十分理想的结果[5]。然而，生成对抗网络的训练相对于其他深度学习网络而言并不稳定，会出现不收敛、梯度消失、模式崩溃等问题[6-8]。因此，本文提出了改进的生成对抗网络模型，以克服训练过程中出现的难收敛、梯度消失、模式崩溃，以及生成数据多样性差的问题。

2.2　模糊测试

模糊测试技术经过20多年的发展已成为一种广泛应用的漏洞挖掘技术。1989年，麦迪逊大学教授 Barton Miller 提出了模糊测试的概念，并测试了 UNIX 系统下应用程序的健壮性。随后，越来越多的研究人员投入到模糊测试的研究中，提出了各种新的思想和方法。例如 Porter 等提出的 PROTOS[9]测试套件使用协议规范生成结构化测试数据，将模糊测试首次应用到网络协议测试中。随后 Aitel 开发了首个自定义的模糊测试器框架 SPIKE[10]，后续也有研究者对 SPIKE 工具做出了改进。Peach[11]用于文件模糊测试，并通过简单易懂的编写语言和跨平台测试受到安全测试人员的喜爱。安全研究员开发的 AFL[12]是基于覆盖引导的模糊测试工具，极大地影响了模糊测试技术，并持续进行改进。最近的研究尝试将深度学习技术与 AFL 相结合，通过应用序列到序列神经网络模型来增强 AFL 模糊器的效果[13]。尽管模糊测试技术

在漏洞挖掘方面凸显了有效性，但其测试用例生成的随机性和变异过程的复杂性仍然是一大挑战。因此，要找到漏洞，需要耗费大量的时间和计算资源。模糊测试中的数据生成过程通常是基于一定的规则和算法[14]，无法生成真实世界中复杂的输入数据，因此很难发现存在的潜在漏洞。另外，目前的模糊测试技术依赖安全专家的测试经验。

在过去的研究中，深度学习方法通常被用作辅助技术工具，但现在有研究将其作为核心方法应用于工业网络协议的模糊测试。例如，Lv 等提出使用机器学习来生成高价值的二进制种子文件[15]。Böttinger 等也提出使用 Q-learning 算法来确定对输入执行的变异操作[16]。Godefroid 等研究了如何利用基于神经网络的学习技术来学习非二进制 PDF 数据对象的语法[17]。这些研究都从不同的角度增强机器学习在模糊测试中的应用。尽管如此，模糊测试主要应用于已知网络协议，对于未知领域的应用并不太适用，这限制了模糊测试在更广泛场景中的应用和发展。

因此，本文使用改进的对抗网络模型，实现对报文结构的深入挖掘与分析，生成大量测试用例；根据对协议报文字段以及漏洞触发原理的分析，提出基于变异算子库和数据包规则的变异策略与方法来指导有效测试用例的生成；最后，通过对测试结果的分析，并在此基础上动态调整变异策略对工控协议进行模糊测试，实现涉及面更深、更广的测试，挖掘出更多的潜在漏洞。生成对抗网络模型与变异策略结合的方法将基于生成与基于变异的模糊测试技术相结合，不仅实现了对未知网络协议结构的识别，扩大了适用范围，同时减少了投入的人工成本与时间成本，而且缩短了找到漏洞所需的时间。

3 基于生成对抗网络与变异策略结合的挖掘方法

3.1 整体架构

本文采用基于生成对抗网络与变异策略结合的漏洞挖掘方法来解决当前网络协议漏洞挖掘面临的难题。其中主要通过提升测试用例的多样性、测试系统接收率，以及漏洞的触发次数等多项评估指标，来缩短生成测试用例的时间，改善模糊测试结果。通过训练改进生成对抗网络，本文方法有效降低了人工分析的时间和成本，消减了人工测试缺乏的客观性，增强了测试用例的多样性。同时，为高效识别出有效用例，本文提出根据已知漏洞特征与协议格式特点引入具有指导性的变异策略与方法。

本文方法的整体架构涉及数据预处理、生成模型设计、变异策略构建以及结果分析与反馈等过程，如图 1 所示。

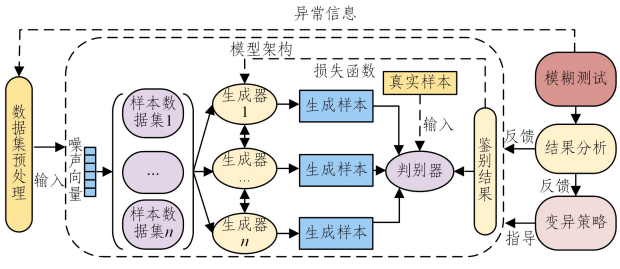

图 1 整体架构图

Fig. 1 Overall framework of the proposed method

首先，对初始报文数据集进行预处理，将处理过的数据输入模型中进行训练优化，实现自动化生成与真实数据结构相似的测试用例。然后结合模糊测试及对测试结果的分析，获取有效的异常信息，识别出有效用例。异常信息主要包括模糊测试过程中的异常响应以及线下漏洞库的内容。最后，将测试结果反馈到数据处理与模型训练中，通过反馈信息对变异策略做出调整，优化对测试用例的变异操作。

3.2 基于图像格式转换的数据预处理

为了进一步增强训练过程，数据预处理流程必须确保将报文数据转换为符合目标格式的图像数据。主要通过对捕获的通讯报文数据集进行数据清洗、进制转换、数据帧对齐、数据聚类以及格式转换操作，将报文数据转化为图像数据，用于模型训练，实现对报文格式的深度学习分析。

工控网络通讯环境下的数据包以序列的形式存在，由报文头部和数据域两部分组成[18]。报文头部包含用于控制数据传输和处理的协议控制信息，具有相对固定的格式。数据域包含应用层的实际传输数据，由于指令信息不同，数据域长度和内容各不相同。值得注意的是，同一协议簇的通讯数据帧通常遵循相似且固定的协议格式。

预处理阶段的主要目标是通过数据帧清洗、对齐、进制转换、格式转换操作把数据处理成模型训练的图片格式。将原始的数据帧序列形式化表示为 $S_{1,n}=(e_1,e_2,e_3,\cdots,e_x,\cdots,e_n)$，$e_x\in E$，$S_{1,n}\in S^*$，其中 E 表示 16 进制的数据集合，由字母和数字构成；S^* 表示数据帧序列集，由数据帧构成。进制转换则将十六进制的协议数据帧转换为十进制数据表示，再存储到特定的文件中。对输入训练模型中的数据格式做统一化处理后，进一步对文件中的数据做格式转换。将一维的向量转换为 3 行 32 列 32 通道的三维数组，“3”表示红、绿、蓝三通道；“32”表示该通道的高度和宽度，即 32 * 32 像素。

数据处理流程如图 2 所示，其中，由于在数据对齐操作过程中加入了 0 填充数据，因此图片 RGB 值为 0 的像素点也很多，训练图片的色调单一。

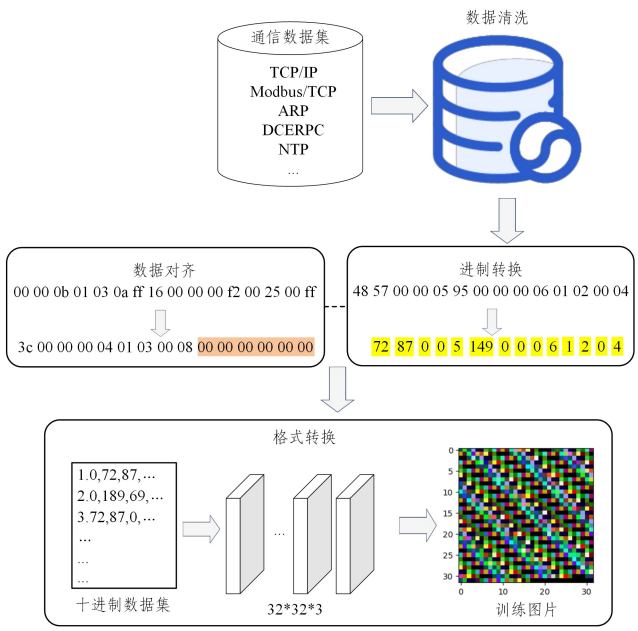

图 2 数据预处理

Fig. 2 Data preprocessing

3.3　改进生成模型设计

基于深度学习中的改进生成对抗网络模型,学习网络协议报文的格式与结构[19-20],生成多样化且具有差异性的测试数据报文,提高模型的泛化能力和鲁棒性。在模型设计过程中,全面考虑模型训练、参数调优、迭代优化等过程,在保证生成样本与实际数据样本之间差异性与相似性的同时,也能够保证生成样本的多样性。

本文提出的改进的生成对抗网络模型(DMGAN)设计原理如下:DMGAN 模型引入了高斯噪声向量,并且运用高斯混合模型(Gaussian Mixture Model,GMM)对隐空间向量进行参数化,再对参数化后的隐空间向量进行重参数化,得到多样化且能反向传播的样本数据;然后从指定的高斯分布中获取样本输入到生成器中,选用多模式生成器学习输入数据报文的空间结构;最后通过辨别器鉴别生成数据的真伪性,当判定结果达到纳什均衡时,结束训练。在对抗网络中,纳什均衡指生成器和判别器都达到最优状态[21],此时生成器生成的样本与真实数据无法被判别器区分开来。这个状态不是静态的,会基于过去的表现对生成器和判别器进行更新和调整。

GMM 是在不增加模型深度的情况下,学习混合模型的参数来增加先验分布的建模能力与生成样本的多样性。通过引入多样性的机制来增强模型的生成能力,使得模型能够在有限数据的情况下生成多样性的样本。其中,高斯分布定义如下:

$$f(x|\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

把 z 的分布定义为一个混合高斯模型。

$$p_z=\sum_{i=1}^{N}\frac{g(z|\mu_i,\sum_i)}{N} \tag{2}$$

其中,$g(z|\mu_i,\sum_i)$ 表示在高斯分布 $N(\mu_i,\sum_i)$ 中取到样本 z 的概率,每个高斯分布都有 μ_i 和 σ_i 两个参数,μ_i 表示均值,σ_i 表示方差。多个高斯分布按照一定比例加权混合,得到新的概率分布。

由于 z 是一个不可微分的随机变量,上述模型的两个参数的梯度不能直接通过样本参数 z 来反向传播,因此引入重参数化技巧。其原理是把 z 转变为可微分变量,即将每个高斯分布进行标准化表示,形式如下:

$$z=\mu_i'+\sigma_i'\epsilon,\epsilon\sim N(0,1) \tag{3}$$

其中,μ_i' 和 σ_i' 是标准化后的重参数,分别表示均值和方差;ϵ 表示辅助噪声变量,即标准正态分布。如此,z 转化成可微分变量。下面给定多个高斯混合分布 $\boldsymbol{p}_z^i$ 为高斯噪声向量:

$$\boldsymbol{p}_z^i=\sum_{j=1}^{N}\frac{g((\mu_i+\sigma_i\epsilon|\mu_j^i,\sum_j^i)}{N} \tag{4}$$

其中,μ_j^i 表示第 i 个高斯混合分布的均值,$\sum_j^i$ 表示第 i 个高斯混合分布的方差。

多模式生成器学习输入数据序列的空间结构,进而生成测试用例,其损失函数由整体损失和局部损失组成。下式为模型训练优化损失函数:

$$\min_{G_{1:K},C}\max_{D}J(G_{1:K},C,D)=E_X\sim P_{\mathrm{data}}[\log D(X)]+E_X\sim P_{\mathrm{model}}[\log(1-D(X))]-\beta\{\sum_{k=1}^{K}\pi_k E_X\sim P_{G_k}[\log C_k(X)]\} \tag{5}$$

式(5)前两项表示多个生成器与辨别器间的损失计算,P_{data} 表示真实数据的分布,P_G 表示生成数据的分布,D 表示辨别器得到的结果。在本文中,多个生成器所得到的分布依次与辨别器计算得到损失值,之后采用求和平均的方式得到生成器与辨别器的整体损失,表示为:

$$D(X)=\frac{P_{\mathrm{data}}(x)}{P_{\mathrm{data}}(x)+P_{G_{1,k}}(x)} \tag{6}$$

式(5)最后一项表示多个生成器之间的损失,公式表示为:

$$L(G_{1:K})=E_X\sim P_{\mathrm{data}}\left[\log\frac{P_{\mathrm{data}}(x)}{P_{\mathrm{data}}(x)+P_{\mathrm{model}}(x)}\right]+E_X\sim P_{\mathrm{model}}\left[\log\frac{P_{\mathrm{model}}(x)}{P_{\mathrm{data}}(x)+P_{\mathrm{model}}(x)}\right]-\beta\{\sum_{k=1}^{K}\pi_k E_X\sim P_{G_k}[\log\sum_{j=1}^{k}\pi_j P_G j(x)]\} \tag{7}$$

本文应用的 DMGAN 模型设置了多模式生成器和鉴别器,并且随机地抽取高斯混合模型重参数化后的样本,然后输入多模式生成器中进行训练。

高斯混合模型中定义 $\boldsymbol{\mu}=[\mu_1,\mu_2,\cdots,\mu_N]^{\mathrm{T}}$ 和 $\boldsymbol{\sigma}=[\sigma_1,\sigma_2,\cdots,\sigma_N]^{\mathrm{T}}$,设置中包含一些简化集合以及每个分量的对角协方差矩阵、等权混合分量等,这限制了模型逼近更复杂分布的能力。并且,通过不断优化调整高斯混合模型参数以及这些参数的比例和权值,更好地拟合潜在数据分布,使得最终生成器产生的 $P_{\mathrm{data}}(G(\mu_i'+\sigma_i'\epsilon)|\epsilon)$ 概率值最大。

DMGAN 模型结构如图 3 所示。将高斯混合模型与多生成器模式相结合,能拟合更为复杂的样本分布,保证生成数据的多样性和有效性,降低模式崩溃与难收敛等问题出现的可能性。

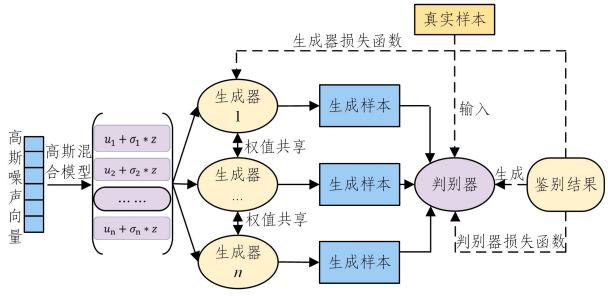

图 3　DMGAN 模型

Fig. 3　DMGAN model

同时设置模型训练参数、学习率、边界裁剪权重以及迭代训练周期等参数。学习率设置为 0.001,权重裁剪值设置为 $[-0.01,0.01]$,设置 500 个迭代周期。高斯混合模型中 μ_i 的取值设定为均匀分布中(−1,1)中的随机值;σ_i 的取值设置为固定值 0.2,需要注意为避免模式崩溃其值不能为 0。将生成器生成的数据与初始数据输入鉴别器中,鉴别器对于两者的区分结果概率均保持在约 1/2 时,即判定已达到稳定的模型训练状态,结束训练。依据设置的 500 个训练周期,保存每 100 个训练时期的生成器模型。通过该方式不仅能获取最终的训练模型,还能保持数据的差异性,生成具有不同相似度的测试用例,提高生成数据的多样性。

训练过程完成后,验证生成器是否可以生成与真实数据具有高度相似性的数据帧。抽取预留的用于模型验证的真实

样本数据集且与生成数据集一同输入鉴别器中进行辨别，确定辨别器已达到纳什均衡状态。

3.4 基于报文算子库的变异策略

变异策略主要用来确定变异的位置，变异方法决定变异的内容，两者共同作用于报文序列，以实现更好的变异效果，确保生成高接收率且种类丰富的测试用例。本文根据现有漏洞规则的已知特征，建立有导向性的报文变异算子库[22]，指导变异策略与变异方法，在可控范围内高效、快速地生成有效用例，提高模糊测试的效率。

对漏洞库进行分析发现，在工控场景下，通讯异常报文会引发地址越界、格式化字符串、缓冲区溢出、整数溢出、逻辑错误以及拒绝服务等问题[23]。这些漏洞通常是程序或设备未对报文中的关键字段做逻辑处理、边界值未经处理、长度过长以及存在特殊字符等原因造成的。例如：缓冲区溢出漏洞的存在是因为未对传入数据帧长度边界值做检查，空指针漏洞的存在是因为未对空做判断处理操作，地址越界的存在也是因为没有对数据帧内容做边界值检查等。

本文针对引发漏洞的异常数据帧集合、异常数据帧特征来设定变异方法，通过改变报文的长度特征、报文中涉及的特殊字符来设定变异方法。对于报文特征字段的变异，我们对字段间的相似性匹配以及报文段大小的边界值做变异。其中，根据已知的协议字段特点定义的初始变异算子如表 1 所列。

表 1 变异算子库

Table 1 Mutation operator library

协议字段	变异算子
Transacation ID	特殊符号，特殊的 ASCII 码
Length	大于准确长度，小于准确长度，非法长度，特殊边界值
Unit ID	合法但未被定义的标识，非法标识，边界值
FC	非法功能码，合法但被设备未定义的功能码，随机字符
Data	超长字符串，单字符、空值，非法读取值，非法数据地址，随机字符，空格，分隔符(! # $ &?,…)，格式化字符串(%d,%n,%x,%f,…)
其他类型	特殊的 ASCII 码，目录遍历符

鉴于现有漏洞的触发原理和协议字段的特点，提出基于变异算子库和数据包规则的变异策略和方法，建立已知触发漏洞的报文特征库并设计变异策略对报文不同区域做出不同的变异操作，以快速地找出有效用例。

3.5 结果分析与反馈调优

将模型训练生成的测试用例与变异生成的测试用例输入测试环境中进行测试。其中，为保证测试接收率和数据用例多样性，先对协议进行结构化分析。设置满足工控通讯特点的模糊测试场景，并将测试数据帧注入设备或者模拟程序进行测试。对接测试程序或测试设备的接口，向其逐步注入测试数据帧，记录测试程序或者测试设备的运转情况，并标记异常通讯数据帧。

为了深入模糊测试，本文对漏洞报文特点进行分析判断，并根据不同特点对报文的变异策略[24]和方法进行调整。同时，分析异常响应的报文，当模糊测试结果中引发某类异常响应的报文比较单一时，动态地更换变异方法以提高测试效果。在确定引发某一异常响应的字段位置后，即对该位置的变异做出约束，以保证变异结果能够引发目标异常响应。同时，对其他位置仍然进行变异操作，以尽可能地发现其他异常情况。综上所述，通过对异常报文特点的分析和判断，并根据反馈的响应信息，动态地调整变异策略和方法，以及对特定字段位置进行约束和对其他位置进行变异操作，可以提高模糊测试的效率和准确性，有效地发现漏洞。

为增强测试用例的多样性，采用变异与模型生成混合迭代的方式筛选有效用例，可以进一步增强变异效果并评估出漏洞报文的特点。优化过程如图 4 所示。

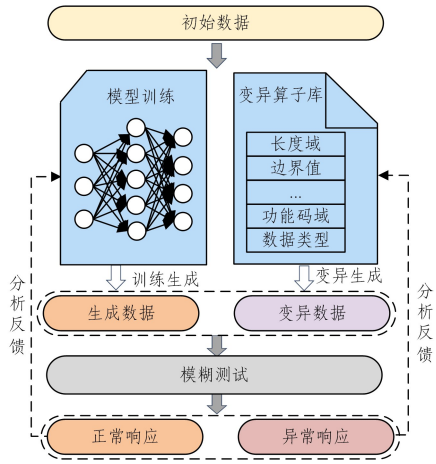

图 4 反馈调优流程

Fig. 4 Feedback process

4 实验

4.1 实验设置

为验证所提方法在各方面指标上呈现出的有效性，搭建实验环境并进行模糊测试来阐明提出的技术亮点。实验中主要针对常用工控网络协议 Modbus-TCP[25-27]及 S7 的通信过程[28]进行模拟仿真测试，来验证测试效果。通过搭建遵循 Modbus 协议规范的通讯仿真工具 Modbus Poll v6.0.2 与 Modbus Slave v6.0.2 和 Modbus RSSim v8.20 以及模拟的串口工具 Configure Virtual Serial Port Driver(VSPD)对 Modbus-TCP 协议进行模糊测试。安装西门子公司推出的 S7-PLSCIM-Advanced V3.0 高功能仿真模拟软件，并配置 PG/PC 接口、PLC IP 地址等完成对 S7 协议通信过程的模拟。最后，采用 Wireshark 工具实现对主站与从站通信过程的异常监测。

本实验采用 Lemay[29]提供的工控数据集来训练，此数据集包括完整流量包捕获和包含恶意流量标签的文件，提供有关数据集生成的详细信息。另外，S7 协议数据集通过模拟环境捕获的数据进行训练。针对模型训练设置，将 DMGAN 模型放在具有 16 个处理器的(Intel(R) Core(TM) i9-12900K CPU@3.20GHz) 32.0GB 内存(RAM) Nvidia GeForce GTX 3090 Ti(24GB) 和 64 位 Microsoft Windows 10 Professional 的机器上训练。与以往基于 WGAN 模型与 WGAN-GP 模型训练的效果进行对比，DMGAN 模型生成的模糊测试用例在各个维度上凸显出了更好的效果。

4.2 实验评估

为体现所提方法的优势，我们引入测试用例的接收率、漏洞的检测能力与生成数据多样性等客观评估指标评估整体

方法框架的有效性。同时，通过生成测试数据的能力与生成数据的多样性来评估改进的对抗网络模型的优势。

通信数据格式不正确，接收方会拒绝接受数据，格式正确则接收方接收。测试接收率指测试目标接收的测试用例百分比，反映生成测试用例的有效性。其定义如下：

$$TA=\frac{n_a}{n_s}\times 100\% \tag{8}$$

其中，n_s是发送用例总数，n_a是接收用例总数。在模型训练与变异过程中，通过调整模型训练参数与变异策略获得更高的测试接收率。

在漏洞挖掘中，通过较少的测试用例挖掘出更多的漏洞是总体实验目标。漏洞检测率能反映发现漏洞的能力，因此将该指标用作衡量方法有效性的最直接标准。其定义如下：

$$AD=\frac{n_b}{n_c}\times 100\% \tag{9}$$

其中，n_b表示有效用例数量，n_c表示测试用例数量。为体现指标的阶段变化性，统计在每 100 个测试用例中通过非正式类比发现的错误数据报文量。该指标是评价目标程序或设备漏洞关联性的最强指标。

引入每小时生成测试用例数量的指标，即模型每小时可以生成的测试用例数量，公式如下：

$$TCGPH=\frac{n_{gc}}{Hours}\times 100\% \tag{10}$$

其中，n_{gc}表示生成用例的数量，Hours 表示生成用例所用的时间。

为体现所提方法的测试范围与能力，将生成数据多样性作为一项重要指标。该指标侧重于生成数据中的类型数量，并展示生成数据多样性的能力。当生成数据类型的数量小于训练数据类型的数量时，可以推断出模型性能不佳，需要进行调优。因此，将其作为模型的训练指标，同时，也可以将该指标作为测试用例的选择标准。

4.3　实验结果

表 2 列出了本文方法与其他方法的对比结果。

表 2　实验结果

Table 2　Experimental results

方法	用例数量	测试目标	测试接收率/%	漏洞检测率/%	漏洞触发次数
DMGAN	26000	Modbus Slave v4.3.4	89.4	6.00	298
		Modbus RSSim v8.20		3.02	187
WGAN	26 000	Modbus Slave v4.3.4	74.3	4.00	97
		Modbus RSSim v8.20		2.61	63
WGAN-GP	26 000	Modbus Slave v4.3.4	88.2	5.57	111
		Modbus RSSim v8.20		4.13	46
Peach 变异	26000	Modbus Slave v4.3.4	48.1	—	18
		Modbus RSSim v8.20			23
DMGAN 结合变异	26000	Modbus Slave v4.3.4	88.5	6.87	329
		Modbus RSSim v8.20		5.92	264

本实验采用 3 种模型与本文模型进行效果比较：基于 WGAN 的模型、基于 WGAN-GP 的模型以及基于 DMGAN 模型，同时对比 peach 工具变异方法。实验将各种方法生成的 26000 个测试数据报文发送到同一配置的模拟从站 Modbus Slave，并观测其通信效果。相比基于 WGAN 以及基于 WGAN-GP 的模型，基于 DMGAN 的模型具有更高的测试接收率，能够触发出更多的异常，其生成用例具有更大的差异性。由于传统变异方法基于对已知字段的变异，且其变异方法和字段均具有随机性，因此其测试接收率较低，同时其漏洞检测率也未被作为参考指标。

图 5 给出了在模型训练周期内的 TA 结果，随着训练时间的增加，TA 上升，这表明有更多生成数据具有正确的消息格式。与基于 WGAN 以及 WGAN-GP 的模型相比，基于 DMGAN 模型的测试接收率在稳定阶段可以达到 89.4%，这表明基于 DMGAN 模型与基于 WGAN 以及 WGAN-GP 模型相比有更高的格式精度。虽然模型进行了不断的调整，但是部分数据格式仍然不正确。最初，测试接收率显著增加；随着训练的不断迭代，测试接收率缓慢增加最后达到稳定。

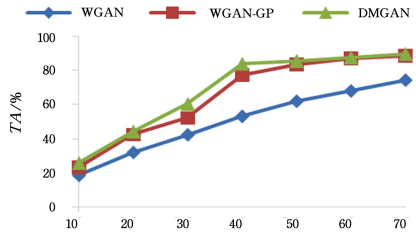

图 5　测试接收率

Fig. 5　Test reception rate

图 6 给出了各模型在模型训练周期内的漏洞检测能力。可以看出，本文方法随着训练时间的增加，检测率都有提升，异常通信的数量也越来越多，最终达到稳定的平坦阶段。本文方法在实验中达到的水平不仅与实验方法有关，也与测试目标有关。实验中选择 Modbus Slave 作为测试对象，基于改进的 DMGAN 模型发现错误的能力比基于 WGAN 与 WGAN-GP 的模型更强，这验证了所提方法的有效性和潜力。

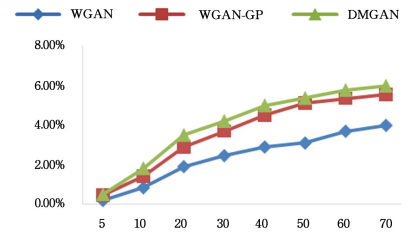

图 6　漏洞检测率

Fig. 6　Vulnerability detection rate

在用例生成时间方面，我们通过实验先后对 DMGAN，WGAN 与 WGAN-GP 模型的效率进行了验证，发现在生成相同数量的用例时，DMGAN 模型生成时间更短，其结果如图 7 所示。其中，epoch 都设置为 30 000 次，每次比较的生成数据集都相同，生成用例个数设定为 1 000，5 000，10 000。

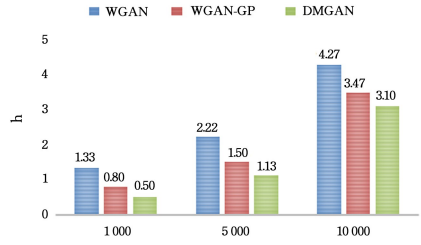

图 7　例生成时间

Fig. 7　Case generation time

经分析比较，原始训练数据中的数据类型在经过 WGAN 模型与 DMGAN 模型训练后都保持了原始数据的多样性，DMGAN 模型产生的数据存在类间差异性。因此，基于 DMGAN 的模型在保持数据的多样性方面比基于 WGAN 的模型更具优势。通常数据类型越丰富，检测异常的能力就越强。因此，基于 DMGAN 的模型可以检测更多错误，图 8 给出了生成数据集基于 *K*-Means 方法的聚类效果。

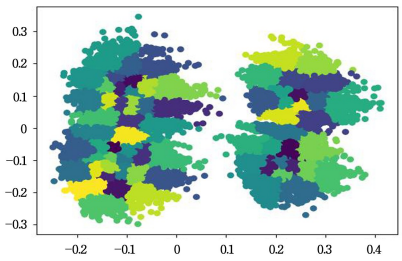

图 8 生成数据聚类效果

Fig. 8 *K*-means performance of generated data

表 3 列出了通信测试中出现的一些异常情况。表中第一列表示本文方法触发的协议异常特征描述。从表中可以看出，本文提出的模糊测试方法不仅可以保证发现漏洞的能力，而且提高了发现漏洞的频率，测试效率更高。

表 3 异常结果

Table 3 Exceptions results

Exceptions	WGAN	WGAN-GP	DMGAN
Slave Crash	16	37	179
Station ID XX Off-Line	57	53	49
Software Caused Connection abort	23	19	23
Integer overflow	71	81	81
File not found	26	30	30
Illegal data address	119	135	169
Invalid Initialization	21	36	26
Illegal Function Code	160	197	297
Writer/Read Error	101	159	206

下面详细描述其中的错误。当测试用例攻击 Modbus_Rssim 时会导致崩溃。在发送大约 1400 条数据帧时，会弹出程序崩溃的提示框。再次对这些数据帧进行通信测试，将其发送到 Modbus Slave。然而，没有异常发生。这表明 Modbus_Rssim 在其实施中存在缺陷。其中，在测试 Modbus Slave 时出现了“write Error”“Read Error”错误后仍然执行对应正确的指令操作，这是仿真程序没有读写功能的读写操作导致的，但这可以显示出本文方法揭示软件错误的能力。在进一步漏洞挖掘过程中发现，Modbus_Rssim 提示异常信息，显示消息“站号 XX 下线，无响应发送”。测试软件有时会断开连接，通过分析发现，这是内存溢出导致的软件崩溃，表明设计实现仿真器时未充分考虑到数据边界填充的情况。在仿真环境进一步的测试中，还发现了“功能码异常”“数据长度不匹配”“整数溢出”和“地址异常”等情况。由于这些异常具有普遍性且在之前的研究中也进行了说明，因此只记录了触发这些异常的测试用例，并提供给模型再次训练，以确定引发异常的报文格式。通过对异常的分析可知，相同的异常行为可能是由不同的原因引起的，然而，不同异常行为也可能是由相同的原因引起的。在模拟实验中，由于测试目标没有源代码，因此未能进一步确定异常背后的具体原理。

最后，将该模型应用在 Simens S7 工控漏洞挖掘中，遗憾的是，Simens S7 只进行发收包，并未检测模型通信过程的异常。但是，在实验过程中能够进行发收包和监控说明了该方法的可行性。

综上所述，改良后的模型生成的测试用例无论是在生成有效性方面还是真正能够检测出漏洞的能力方面都优于改良前的模型。随着训练次数的增加，生成对抗模型生成的有效测试用例占比增加，可以引发漏洞的协议数据帧也增加，证实了该方法的可行性。总体而言，该方法能够实现对未知协议的漏洞挖掘，且效果比较理想，对于复杂的工控协议则需要长时间的训练才能成功。

结束语 本文采用基于生成对抗网络与变异策略结合的网络协议漏洞挖掘方法，提出了一种新的生成对抗网络模型(DMGAN)，并将该模型与线下快速变异策略结合应用于工控网络协议的漏洞挖掘中，实现了在无人工分析的情况下，智能化分析和学习网络通信过程中的漏洞数据报文的格式，更加快速高效地挖掘出网络协议的漏洞。同时，该方法能够适应未知网络协议的漏洞挖掘过程。

本文改进的模型一定程度上解决了训练过程中模式崩溃的问题，适用于更少量的数据集的场景，但是该模型仍然存在一定的缺陷，如训练时间长，训练效果依赖于训练数据集的质量；线下变异策略也没有实现模糊测试完整流程上的完全智能化，需要动态匹配调整。未来可以考虑对漏洞规则进行学习，建立对漏洞规则的完整记忆和特征记忆，实现更加高效快速的智能化漏洞挖掘。

参 考 文 献

[1] LI D. Analysis of the Earthquake Network Virus Event and Enlightenment on Improving Industrial Control Security Protection Capability [J]. Network Security Technology and Application, 2019, 217(1): 9-10, 24.

[2] RUI X. 2020 China Network Security Report [J]. Research on Information Security, 2021, 7(2): 102-109.

[3] WHITEHEAD D E, OWENS K, GAMMEL D, et al. Ukraine cyber-induced power outage: Analysis and practical mitigation strategies[C]// 2017 70th Annual Conference for Protective Relay Engineers(CPRE). IEEE, 2017: 1-8.

[4] KURDS J F, WROSS K. Computer Networking A Top-Down Approach Seventh Edition[M]. China Machine Press, 2021.

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[C]// Neural Information Processing Systems. MIT Press, 2014.

[6] WANG Z L, ZHANG B W. Overview of Research on Generative Adversarial Networks [J]. Journal of Network and Information Security, 2021, 7(4): 68-85.

[7] GURUMURTHY S, SARVADEVABHATLA R K, RADHAKRISHNAN V B. DeLiGAN: Generative Adversarial Networks for Diverse and Limited Data[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2017.

[8] GHOSH A, KULHARIA V, NAMBOODIRI V, et al. Multi-Agent Diverse Generative Adversarial Networks [J]. arXiv:

1706.02906，2017.

[9] PORTER B W，BAREISS E R. PROTOS：An Experiment in Knowledge Acquisition for Heuristic Classification Tasks[M]. University of Texas at Austin，1986.

[10] AITEL D. MSRPC Fuzzing with SPIKE 2006[J/OL]. http://www.immunitysec/spike.html.

[11] KIM M，PARK S，YOON J，et al. File Analysis Data Auto-Creation Model For Peach Fuzzing[J]. Journal of the Korea Institute of Information Security and Cryptology，2014，24(2)：327-333.

[12] JI T，WANG Z，TIAN Z，et al. AFLPro：Direction sensitive fuzzing[J]. Journal of Information Security and Applications，2020，54：102497.

[13] ZALEWSKI M. Americanfuzzylop.[EB/OL]. http://lcamtuf.c.cx/aflfl/.

[14] LAI Y X，YANG K X，LIU J，et al. Mining Method for Industrial Control Network Protocol Vulnerability Based on Fuzzy Testing [J]. Computer Integrated Manufacturing System，2019，25(9)：2265-2279.

[15] LV C，JI S，LI Y，et al. SmartSeed：Smart Seed Generation for Efficient Fuzzing[J]. arXiv：1807.02606，2018.

[16] BOTTINGER K，GODEFROID P，SINGH R. Deep Reinforcement Fuzzing[C]// 2018 IEEE Security and Privacy Workshops (SPW). IEEE，2018.

[17] GODEFROID P，SINGH R，PELEG H. Machine Learning for Input Fuzzing：US patent，20180285186A1[P]. 2018.

[18] ZHAO H，LI Z，WEI H，et al. SeqFuzzer：An Industrial Protocol Fuzzing Framework from a Deep Learning Perspective[C]// IEEE Conference on Software Testing，Validation and Verification. East China Normal University，2019.

[19] LIN P Y，TIEN C W，HUANG T C，et al. ICPFuzzer：proprietary communication protocol fuzzing by using machine learning and feedback strategies[J]. Cybersecurity，2021，4(1)：1-15.

[20] SONG C X，YU B，ZHOU X，et al. SPFuzz：A Hierarchical Scheduling Framework For Stateful Network Protocol Fuzzing [J]. IEEE Access，2019，7：18490-18499.

[21] LI Z，ZHAO H，SHI J，et al. An Intelligent Fuzzing Data Generation Method Based on Deep Adversarial Learning[J]. IEEE Access，2019，7：49327-49340.

[22] PANT M，ALI M，ABRAHAM A. Mixed mutation strategy embedded differential evolution[C]// 2009 IEEE Congress on Evolutionary Computation. IEEE，2009：1240-1246.

[23] LI W M，ZHANG A F，LIU J C，et al. Automated Fuzzy Testing Vulnerability Mining Method for Network Protocol [J]. Chinese Journal of Computer，2011，34(2)：242-255.

[24] DENG J，ZHU X，XIAO X，et al. Fuzzing With Optimized Grammar-Aware Mutation Strategies [J]. IEEE Access，2021，9：95061-95071.

[25] LAI Y，GAO H，LIU J. Vulnerability Mining Method for the Modbus TCP Using an Anti-Sample Fuzzer[J]. Sensors，2020，20(7)：2040.

[26] SASI A，HARIPRASAD K V，CHERIAN S，et al. R0fuzz：A Collaborative Fuzzer for ICS Protocols[C]// 2021 12th International Conference on Computing Communication and Networking Technologies(ICCCNT). 2021.

[27] XU Y，YI Y，LI T，et al. Review on cyber vulnerabilities of communication protocols in industrial control systems[C]// 2017 IEEE Conference on Energy Internet and Energy System Integration(EI2). IEEE，2017.

[28] HU Z，SHI J，HUANG Y H，et al. GANFuzz：a GAN-based industrial network protocol fuzzing framework[C]// the 15th ACM International Conference. ACM，2018.

[29] LEMAY A，FERNANDEZ J M. Providing {SCADA} Network Data Sets for Intrusion Detection Research [C]// 9th Workshop on Cyber Security Experimentation and Test ({CSET} 16). 2016.

ZHUANG Yuan，born in 1988，Ph.D，lecturer，associate professor，master's supervisor. Her main research interests include blockchain security，machine learning，big data processing and distributed computing.

SHEN Linshan，born in 1978，master，associate professor，master's supervisor. His main research interests include industrial information security，machine learning and intelligent information processing.

（责任编辑：何杨）