基于生成对抗网络的网络安全漏洞检测技术研究

文◆中国电子科技集团公司第三十研究所　杨伟斌　吕树林　张红珍　叶子琛

引言

在互联网时代下，网络安全漏洞问题日益凸显，根据有关数据，截至 2024 年，全球因软件漏洞导致的经济损失超 1.2 万亿美元 [1]。传统的网络漏洞检测技术主要以分析测试、渗透测试等为主，存在覆盖率较小、人工依赖等短板，难以有效满足网络安全漏洞检测要求 [2]。为此，本文针对传统检测技术的不足，提出一种基于生成对抗网络的网络安全漏洞检测方法，通过生成对抗网络模型，挖掘报文序列中的信息，提取相关特征，并通过基于变异算子库指导的迭代变异策略，构建测试生成规则。测试结果表明，在训练时间不断增加的情况下，本文方法的 TA（不同方法接受率）始终保持在最大数值，在稳定阶段可达 90.01%，且漏洞检测率较高，该方法能够有效解决网络安全漏洞检测的实际需求。

1 基于生成对抗网络的网络安全漏洞检测方法

1.1 整体结构

本方法整体结构如图 1 所示。

（1）第一步，对网络报文数据进行收集，在完成数据预处理后，将数据输入模型，并生成相应的测试用例。

（2）第二步，基于模糊测试结果，挖掘有效异常信息（包括异常响应、线下漏洞库的内容），并对其中有效用例进行识别 [3]。

（3）第三步，将测试结果反馈到数据处理与模型训练，并通过变异策略进行调整。

1.2 数据预处理

在网络环境中，数据包主要为序列形式，包括报文头部与数据域。为有效提升模型的训练效果，需对报文数据进行处理，包括数据帧清洗、对齐、进制转换等。

（1）数据帧清洗。去除噪声

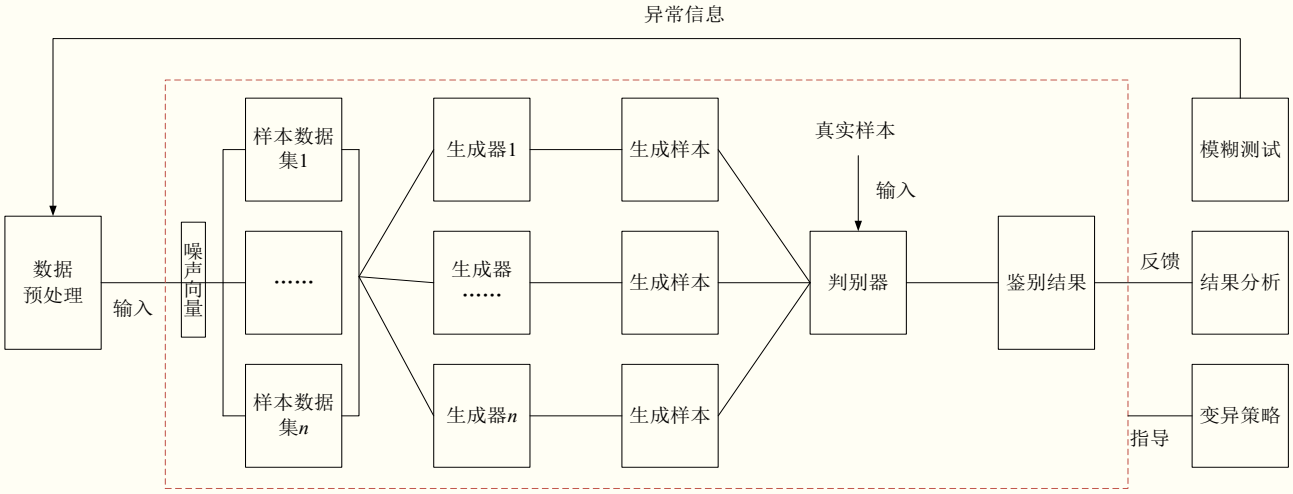

图 1 整体结构

【作者简介】杨伟斌（1991—），男，甘肃定西人，本科，工程师，研究方向：网络安全。

数据（如损坏的帧、重复的包）、填充缺失字段或纠正错误格式。

（2）数据对齐。统一报文长度（如填充或截断），确保输入模型的序列长度一致。

（3）进制转换。将原始二进制数据转换为模型可处理的数值形式（如十进制整数或浮点数）或直接保留为二进制流（通过嵌入层处理）。

1.3 模型改进

为有效提升生成对抗网络模型的泛化能力的鲁棒性，本研究结合深度学习算法，对模型进行相应的改进，具体步骤如下。

第一步，通过在模型中添加高斯噪声，并借助高斯混合模型（Gaussian Mixture Model，GMM）对低维向量进行参数化建模，进而利用重参数化技术获得多样化样本。

第二步，在对应的高斯分布中提取样本，基于生成器输入相应的数据空间结构。

第三步，借助辨别器分析数据真实性，当结果达到纳什均衡，则结束训练。

对于GMM模型而言，其主要是在学习混合模型参数的基础上引入多样性机制。在整个过程中，高斯分布表示如式（1）所示。

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

假设z为一个混合高斯模型，则有式（2）。

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

式（1）（2）中，$g(z|\mu_i,\sum_i)$为高斯分布中取到z的概率；μ_i、σ_i为参数，分别表示均值与方差。根据一定的比例，对高斯分布进行加权混合后，便可获取相应的概率分布。

其中，z属于随机变量，具有不可微分的特点。为此，对z进行重参数化处理，则有式（3）。

$$z=\mu_i^{'}+\sigma_{i\in,\epsilon}^{'}\sim N(0,1) \quad (3)$$

式（3）中，μ'_i、σ'_i分别为经过重参数化的均值与方差；$\in$为辅助噪声变量。在经过相应的转换后，z便可进行微分。

此时，多个高斯混合分布p_z^i可表示为式（4）。

$$p_z^i=\sum_{j=1}^{N}\frac{g(\mu_i+\sigma_i\in|\mu_j^i)}{N} \quad (4)$$

对于改进生成对抗网络模型（DMGAN），其生成器与鉴别器包含多个模式，通过将高斯混合模型与此种模式进行结合后，可在一定程度上提升数据的多样性与有效性，避免出现模型难以收敛的问题。

1.4 变异策略

对于变异策略而言，其作用在于识别变异位置。本研究中，在现有网络漏洞的基础上，构建相应的报文变异算子库，并基于变异策略，有效提升模糊测试的效率。在通讯异常的情况下，漏洞库存在格式化字符串、缓冲区溢出等问题。针对这一系列问题，首先通过分析报文长度和字段特征的变换规律建立变异模型，其中结合协议字段的相似性匹配和尺寸特征来设计初始变异算子。其次，根据协议特性和漏洞触发机制制定自适应变异策略，构建已知漏洞报文特征数据库。最后，采用分区变异机制生成测试用例。系统具备动态调节能力，可根据漏洞报文特征实时调整变异参数，并分析异常响应。当发现异常报文多样性不足时，将精确定位目标字段并实施变异约束，同时维持其他字段的变异强度以确保测试覆盖率。

2 实验分析

2.1 实验环境搭建

为验证本文方法的测试效果，研究通过Modbus TCP网络协议、S7通信过程，构建仿真测试环境。其中，模糊测试主要通过Modbus Poll v6.0.2、Modbus S1ave v6.0.2进行，通信过程主要通过S7–PLSCIM–Advanced V3.0软件模拟，异常检测主要采用wireshark工具完成。在实验过程中，以工控数据作为原始数据进行训练，其中包括完整流量包、恶意流量标签，而S7协议数据集通过模拟环境捕获的数据进行训练。

2.2 评估方法

将本文方法与基于WGAN模型与WGAN GP模型进行对比，选取接受率、漏洞检测率、生成数据多样性作为评价指标，对不同方法的有效性进行评估。

2.3 实验结果分析

2.3.1 不同方法接受率（TA）对比分析

在训练周期内，不同方法测试接受率（TA）对比结果如图2所示。

从图2可看出，在训练时间不断增加的情况下，三种方法的TA均呈现逐渐升高的趋势，说明不同方法生成的数据具有一定的准确性。但是，本文方法的TA始终保持在最大数值，在稳定阶段可达90.01%，说

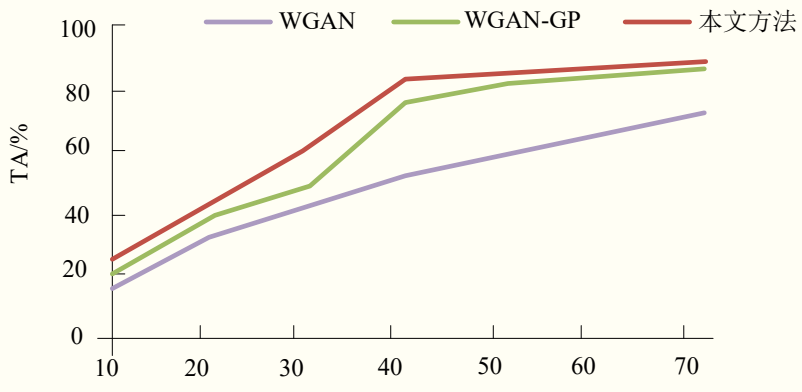

图 2 不同方法测试接受率（TA）对比结果

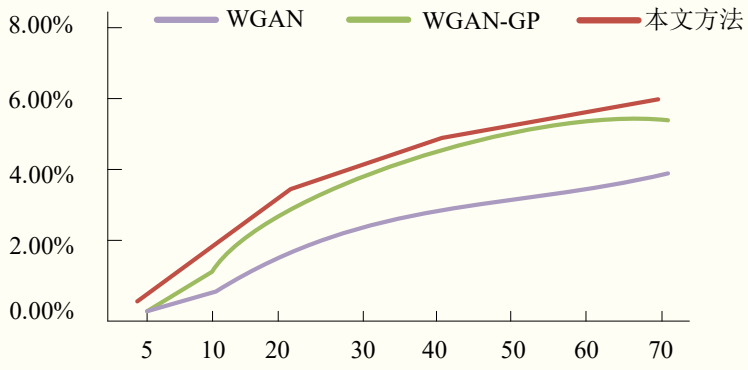

图 3 不同方法漏洞检测能力（AD）对比结果

明本文方法格式精度相对较高。此外，本研究虽然对模型进行了调整，但部分数据格式依然存在相应的问题。总之，在迭代次数不断增加的情况下，各方法的 TA 从最开始的持续增加，逐渐趋近于稳定。

2.3.2 **不同方法漏洞检测能力（AD）分析**

在训练周期内，不同方法漏洞检测能力（AD）对比结果如图 3 所示。

从图 3 中可以明确看出，在训练时间不断增加的情况下，三种方法的 AD 均呈现出一种逐渐升高的趋势，且异常通信数量也逐渐增多，并逐渐趋近于稳定。相较之下，本文方法漏洞检测率较高，其根本原因是测试目标的不同，本文方法的测试对象主要为 ModBus Slave，使本文方法发现错误的能力高于另外两种方法，说明本文方法具有一定的有效性。

2.3.3 **不同方法用例生成时间对比分析**

在用例生成数量相同的情况下，本文方法所需的时间最短。此外，在通信测试过程中，出现了部分异常情况，异常结果中存在的错误表述如下。在 Modbus_Rssim 受到测试用例的攻击时，则会发生软件崩溃。在数据帧条数为 1400 条的情况下，自动弹出崩溃提醒。通过对数据帧格式进行二次测试且发送至 Modbus_Slave 后发现，无异常情况产生，提示 Modbus_Rssim 存在问题。在对其中的漏洞进行进一步挖掘后发现，Modbus_Rssim 提示存在异常信息，且显示“无响应发送”，究其原因主要是内存溢出导致的软件崩溃，说明在仿真过程中忽略了数据边界填充的问题。通过进一步测试得知，其中还存在数据长短不匹配、地址异常等问题。综合分析来看，本文方法具有较高的检测精度，且具有一定的可行性。

结语

本研究针对传统的网络漏洞检测技术存在的不足，提出一种基于生成对抗网络的网络安全漏洞检测方法，并结合变异策略对其应用进行验证。研究结果显示，本方法可在无人工分析的条件下，实现网络漏洞数据的智能分析，进一步提升了网络协议中的漏洞检测效率，对于提升网络安全有着十分重要的作用。但是，本文方法还存在相应的不足，如训练时间长、变异策略无法进行动态调整等。在后续的研究中，还需深入研究漏洞规则，进而提升网络安全漏洞挖掘的智能化水平。

引用

[1] 靳文京,卜哲,秦博阳.基于序列生成对抗网络的智能模糊测试方法[J].信息安全研究,2024,10(6):490-497.

[2] 金磊.面向网络空间防御的越权漏洞对抗机器学习检测系统[J].微型电脑应用,2025,41(2):292-296.

[3] 王梦雨,朱树永,张玉军.一种基于行为特征的网络靶场大规模攻击流量生成方法[J].高技术通讯,2024,34(11):1153-1163.