 DOI: https://doi.org/10.12349/iser.v5i8.3244

Research and Implementation of Advanced Persistent Threat Detection Algorithm Based on Generative Adversarial Network

Yang Gao Wei Chen Xuezhong Lu

State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi, Xinjiang, 830000, China

Abstract

With the continuous evolution of cyber attack means, advanced persistent threat (APT) poses an extremely severe challenge to the entire network security system. This paper proposes an advanced persistent threat detection algorithm based on generative adversarial network (GAN), aiming to significantly improve the accuracy and real-time response ability in the detection process. The basic concept of advanced persistent threat and the detection difficulties in practice are expounded in detail, and the imperfections of the existing detection methods are analyzed. This paper then details the basic principles of generating an adversarial network and its applicable field and potential for advanced persistent threat detection, and how to be used for more effective APT detection. This paper designs a novel detection framework and implements a specific and efficient detection algorithm. Through comprehensive and detailed experimental evaluation, the effectiveness of the proposed algorithm in APT.

Keywords

advanced persistent threat; generating adversarial network; network security; detection algorithm

基于生成对抗网络的高级持续性威胁检测算法研究与实现

高阳 陈伟 鲁学仲

国网新疆电力有限公司信息通信公司，中国 · 新疆 乌鲁木齐 830000

摘 要

随着网络攻击手段的持续演进，高级持续性威胁（APT）对整个网络安全体系构成了极其严峻的挑战。论文提出一种基于生成对抗网络（GAN）的高级持续性威胁检测算法，旨在显著提高检测过程中的准确性和实时响应能力。先对高级持续性威胁的基本概念及其在实际操作中存在的检测难点进行了详细而深入的阐述，同时分析现有检测方法的不完善之处。随后论文详细介绍了生成对抗网络这一技术的基本原理以及它在高级持续性威胁检测中的适用领域和潜力，以及如何利用进行更为有效的APT检测。论文设计一种新颖的检测框架，实现了具体而有效的检测算法。通过全面细致的实验评估，验证确认了所提议算法在实际应用中进行APT检测时表现有效性。

关键词

高级持续性威胁；生成对抗网络；网络安全；检测算法

1 引言

在当今的当前网络环境中，高级持续性威胁（APT）由于其非常高的隐蔽性、极强的持续性以及目标的明确性，对关键信息基础设施构成了严重而巨大的威胁。传统常规的安全检测方法在应对 APT 攻击时，往往因为缺乏足够的智能和灵活适应性，而显得捉襟见肘、难以奏效。生成对抗网络（GAN），作为一种强大有力而先进的深度学习模型，以其在图像生成以及风格迁移等领域中的成功应用，为 APT 检测提供了一种全新的独特视角。论文详细地探讨了 GAN 在 APT 检测中的潜力和应用，提出了一种创新性、基于 GAN 的高效检测算法。该算法通过模拟复杂的攻击者与防御者之间的博弈过程，实现对 APT 行为的高效识别。这一新颖的算法不仅显著提高检测准确性，还能增强系统自适应能力，为网络安全防护提供了一项全新的解决方案。

【作者简介】高阳（1988–），男，中国天津人，本科，高级工程师，从事网络安全技术研究。

2 研究综述

2.1 高级持续性威胁（APT）概述

高级持续威胁（APT）是一种复杂长期的网络攻击方法，通常由训练有素、有组织的团体发动。这些团体拥有明确目标，在目标网络中长时间潜伏，以获取敏感和重要的信息或破坏关键基础设施 [1]。APT 攻击者具备极高技术能力和丰富资源，可以巧妙绕过常规安全防护措施。他们多样而复杂

的攻击手段包括利用零日漏洞、社会工程技巧以及使用高度隐蔽性恶意软件等方式。APT 攻击因其隐蔽性与持续性使得早期难以被察觉，一旦成功渗透，就能对目标系统造成长期深远影响。

2.2 现有 APT 检测方法的局限性

虽然现阶段已有 APT 检测方法，包括基于签名的探测、异常行为评估和网络流量分析等，能够识别并防御网络攻击，但在应对 APT 时依然存在明显短板。具体而言，基于签名的探测依赖已知攻击模式进行匹配，因此难以处理未知或变种形式的威胁。尽管异常行为分析技术能发现与正常活动不符的不寻常情况，但 APT 攻击者通过精心模拟普通用户操作，可大幅降低被侦测到的可能性。而且对于网络流量分析在面对大量复杂数据时，经常出现误报和漏报问题，并且从海量数据中精准识别出 APT 也非常艰难。

2.3 生成对抗网络（GAN）在 APT 检测中的潜力与挑战

生成对抗网络（GAN）是一种新型深度学习方法，通过生成者和判别者的持续竞争训练，展现出显著的模式识别和数据生成能力。在 APT 探测领域，GAN 可以仿真 APT 行为的数据，从而增强检测模型适应性，并通过训练判别器，提高对 APT 行为的精准识别。将 GAN 用于 APT 探测仍面临诸多复杂挑战，包括设计更有效高效的网络结构以捕捉 ATP 攻击复杂特征，需平衡生成者与判别者之间训练，以防止模式崩溃，还必须确保产生的数据能够准确反映 ATP 攻击的一些独有特点。

2.4 研究意义与目标

由于 APT 攻击的极度严重性以及现有检测方法的显著不足，本研究目标是探索一种基于 GAN 的 APT 检测算法，目的是为了大幅提高检测精准性的同时增强实时性能。该研究的重要意义在于通过引入先进且复杂的机器学习技术，致力于提升整个网络安全防护体系的智能化水平和自适应能力，从而实现更高效、更可靠的保护措施。本研究项目的主要目标是设计和实现一种基于 GAN 的 APT 检测框架，这个框架能够自动学习并识别 APT 攻击行为模式，并且在真实的网络环境中有效地识别出并响应这些恶意 APT 攻击。通过对生成对抗网络模型进行深入的研究和优化，本研究期望为广大的网络安全领域提供一个更加高效能的高级持续性威胁检测解决方案。

3 生成对抗网络基础

3.1 GAN 的基本原理

生成对抗网络(GAN),这是由一个叫生成器(Generator ）的部分和另一个被称为判别器（ Discriminator ）的部分组成的复杂深度学习模型。这个模型的核心思想是，通过这两者之间不断进行反复、激烈的对抗，来逐渐优化整个模型的参数。具体来说，生成器的主要目标是尽其所能地去生成那些看起来非常逼真的数据，而判别器则肩负着非常且艰巨的任务即要在真实的数据和由生成器制造出的假数据之间，尽量准确地做出区分。这个过程可以用警察和狡猾伪造者之间不断的智慧博弈来形象地类比，生成器在持续学习中逐步掌握如何制造那些更加难以被识别的高质量伪造品，而判别器则在这一过程中不断提高其精准的识别能力。随着漫长的训练过程，生成器所产生的数据品质日益提升，最终能够达到以假乱真的逼真效果 [2]。GAN 这种独特而出色的特性使得它在图像合成、风格迁移等多个领域中展现出了非常卓越的性能表现，对抗网络训练过程如图 1 所示。

图 1 对抗网络训练过程

3.2 GAN 的关键组件与工作流程

GAN 的整个工作流程主要包括三个的阶段：数据准备、网络训练和模型评估。在最初的数据准备阶段，必须收集大量的真实数据信息供后续训练使用。网络训练阶段涉及生成器和判别器两模块交替进行复杂而密集的训练。生成器通常会采用卷积神经网络（CNN），因为其特别适合捕捉数据的分布特征结构，而判别器则可能使用 CNN 或多层感知机（MLP）这样的灵活架构来认真学习区分真实数据和虚假数据。在对抗训练这个关键过程中，生成器和判别器的损失函数是相互竞争、不断较量的。

3.3 GAN 在安全领域的应用现状

当前网络安全形势复杂而具挑战性。生成对抗网络（GAN）在恶意软件识别、异常流量检测和数据隐私保护方面应用广泛。在恶意程序检测中，GAN 通过创建多样化的恶意代码实例，提高模型普适性，从而增强其精确度和可靠性。在发现异常流量方面，GAN 可以生成大量正常流量数据集，使模型理解常见行为模式，从而提升反常行为辨识准确度。在数据隐私防护领域，通过生成虚拟数据集来保护用户信息，同时为机器学习提供有价值的训练素材。然而，即使生产对抗网络在安全领域展现巨大潜力，它仍面临如数据分布不均、过拟合及训练过程稳定性差等问题。

4 高级持续性威胁检测算法设计

4.1 检测算法框架设计

在这项研究中，建议的 APT 检测算法框架设计以 GAN

为核心构建，围绕其特性建立了完整的检测流程。首要的是该框架通过数据采集模块收集多源数据，包括网络流量和系统日志等各种信息。然后数据预处理模块对原始信息进行清洗和标准化处理，以消除无意义噪音，并提取有用的数据。在此基础上特征工程模块进一步深入挖掘并构造能代表 APT 攻击行为的重要特征。核心的 GAN 模块由生成器和判别器组成。最终在结果分析阶段，检测结果传送到分析模块，该模块根据判别器提供的信息判断 APT 攻击，并及时发出警报信号。

4.2 数据预处理与特征工程

数据的预处理是必不可少的，通过预处理可以让后面的过程事半功倍，可以去除那些不想管的数据，也可以填补一些空缺值，以及让数据尽可能地标准化，以确保数据质量。特征工程在实行时需要深入了解到 ATP 攻击的特点，从最原始的数据那里搜寻有用的关键信息，比如网络连接模式、系统调用频率、用户日志等。使用应用特征选择技术与降维方法来进一步优化这些属性，达到一个提高模型适应能力和减少计算负担的作用。这些关键属性不仅要能准确识别 APT 攻击，还必须具备足够的鲁棒性以应对复杂的攻击。

5 算法性能评估

5.1 评估指标与测试环境

在评估 APT 检测算法实用性的时候，我们需要用到一些综合性的标准，这些标准包括检出率、误判率、漏报率、响应时间和模型泛化能力，通过衡量这些标准的优良，我们可以推断出其在实际应用中的真实效果。另一方面，在评估的过程中，搭建合适的测试环境是至关重要的，只有搭建了合适的测试环境，才能真实反映情况，需要模拟的部分有“真实”网络场景、各类网络流量、多种系统日志和用户行为数据，为了确保覆盖面是广阔且可靠的，所使用的数据还应该要涵盖多样式的 APT 攻击场景 [3]。

5.2 实验设计与实验结果

该实验设计一定要遵循科学性和可重复性的基本原则，通过实验验证提高算法效能。整个实验分为两个环节：第一阶段是模型训练，研究者调试 GAN 的超参数来提升模型架构；第二阶段是模型评估，使用独立测试集测量算法在各方面的性能指标。结果表明，基于 GAN 的 APT 检测方法在准确率和响应时间均上优于传统方法，在泛化能力方面也表现出色，可适应各种不同类型 APT 攻击模式。

5.3 算法性能分析

对实验结果的深入、详细分析揭示了所提算法的显著优势和潜在的改进空间。这种检测准确率高表现主要归因于 GAN 在模拟 APT 攻击行为以及区分真假数据方面展示出的高效能力。对响应时间的优化，实际上是得益于算法实现过程中的效率考量以及并行计算技术的应用。然而误报率和漏报率的详细分析则指出了模型在面对某些特定攻击模式时所表现出的性能局限。这一发现提示我们，在未来研究中需要进一步优化模型的判别能力，以及在特征工程方面进行更多改进。

5.4 与其他方法的比较

为了能够更加全面地评估所提算法的整体性能，可以将其与其他现有的 APT 检测方法进行比较。当处理那些复杂且多变的 APT 攻击时，基于 GAN 的这类检测算法展现出了更强大的适应性和识别能力。尤其是在面对那些未知且新颖的攻击模式时，由 GAN 生成的一系列模拟数据显著增强模型的泛化能力，从而减少了依赖先验知识这一点。此外与其他的机器学习方法相比较， GAN 在自动化特征学习和模式识别领域展现出了明显的优势，为 APT 的检测提供了全新的研究方向和技术手段 [4]。举例 APT 检测方法比较概览如表 1 所示。

表 1 APT 检测方法比较概览

检测方法	核心概念	主要优点	潜在缺点
基于 GAN 的 APT 检测	生成器和判别器的对抗训练	高准确性，适应性强	资源消耗大，训练要求高
传统签名检测	基于已知攻击特征	实时检测快，部署简单	对未知攻击无效
异常行为分析	行为模式分析	适应未知攻击	误报率高，资源消耗
基于启发式的方法	多阶段攻击技术检测	可检测复杂攻击链	依赖先验知识，更新频繁

6 结论

论文提出了一种基于 GAN 的 APT 检测方法，利用具有创新性的架构和对抗训练机制，达到了高效识别 APT 攻击的目的。研究结果显示，该方法在准确率、响应时间及适应能力等方面，相较于传统方法来说具有明显的优势。另一方面，通过类似模拟攻防竞赛的方式，算法能不断学习并适应新的 APT 策略，从而大幅提升自身检测系统的智能化和自适应水平。论文通过深入分析，为优化算法提供清晰的研究方向，同时验证了 GAN 在 APT 检测中的应用前景。

参考文献

[1] 程艳艳,孙滨.面向高级持续性威胁的分布式跨域网络安全监测[J].信息技术与信息化,2024(6):196-200.

[2] 朱靖娴.面向高级持续性威胁的网络安全态势感知方法研究[D].武汉:华中科技大学,2023.

[3] 陈卫平.高级持续性威胁检测与分析技术初探[J].现代电视技术,2018(11):135-137.

[4] 刘嘉,谢冰,杨传旭,等.基于网络行为自学习的高级持续性威胁检测技术研究[J].计算技术与自动化,2019,38(2):108-113.