

北京理工大学

本科生毕业设计(论文)

基于模糊测试的协议脆弱性研究

AFLNetSpy: Greybox Fuzzing of IoT Network Applications

学 院:	网络空间安全学院
专 业:	网络空间安全
班 级:	12112002
学生姓名:	常振轩
学 号:	1120202439
指导教师:	谭毓安

2024 年 5 月 10 日

原创性声明

本人郑重声明：所呈交的毕业设计（论文），是本人在指导老师的指导下独立进行研究所取得的成果。除文中已经注明引用的内容外，本文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

特此申明。

本人签名: _____ 日期: _____ 年 _____ 月 _____ 日

关于使用授权的声明

本人完全了解北京理工大学有关保管、使用毕业设计（论文）的规定，其中包括：①学校有权保管、并向有关部门送交本毕业设计（论文）的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存本毕业设计（论文）；③学校可允许本毕业设计（论文）被查阅或借阅；④学校可以学术交流为目的,复制赠送和交换本毕业设计（论文）；⑤学校可以公布本毕业设计（论文）的全部或部分内容。

本人签名: _____ 日 期: _____ 年 月 日

指导老师签名: _____ 日 期: _____ 年 月 日

基于模糊测试的协议脆弱性研究

摘 要

随着物联网领域的快速发展，基于Linux的固件数量呈爆炸式的同时，也面临着日益严峻的安全问题。在传统的程序测试领域，模糊测试已经被证明是一项十分有用的自动化测试和漏洞挖掘技术。但由于固件程序特殊的执行环境，传统的模糊测试技术难以直接应用于固件程序。目前尚未出现完善的针对固件内应用进行灰盒模糊测试的方法和工具。

本文从模糊测试的角度对协议脆弱性进行研究，提出利用QEMU插件监测客户机系统内的系统调用，并通过页目录地址识别目标测试进程，进而获取目标测试进程代码执行信息，以及利用网络请求探测和控制目标测试进程状态的方法，从而为实现针对固件内网络应用的灰盒模糊测试提供了理论基础。

基于上述方法，本文结合QEMU的AFL-SPY插件，在AFLNet的基础上开发出针对固件网络应用的灰盒模糊测试框架AFLNetSpy。它能够对固件系统内的网络应用程序进行高效的灰盒模糊测试，成功将AFLNet的灰盒模糊测试能力扩展到固件应用领域，对促进固件领域的模糊测试研究具有重要作用。

本文首先介绍研究背景、研究意义以及模糊测试和固件安全等领域的相关工作，之后介绍利用QEMU的AFL-SPY插件获取客户机系统内测试程序代码执行信息和利用网络请求探测和控制目标程序状态的方法。然后介绍AFLNetSpy的系统架构和各模块设计要点，最后设计实验从稳定性和性能两方面对AFLNetSpy进行分析验证。

关键词：灰盒模糊测试；固件安全；网络应用

AFLNetSpy: Greybox Fuzzing of IoT Network Applications

Abstract

With the rapid development of the Internet of Things(IoT) field, the number of Linux-based firmware has grown explosively, and the security issues it faces have become increasingly severe. In the filed of traditional program testing, fuzz testing, or fuzzing, has been proven to be a highly useful technique for automated testing and vulnerability discovering. However, due to the special execution environment, it's difficult to apply traditional fuzzing technique to firmware programs directly. And currently, there are no mature methods or tools available to conduct grey-box fuzz testing for applications within firmware.

In this thesis, we does some research on protocol vulnerability from the perspective of fuzz testing. And we propose a way using a QEMU plugin to monitor system calls within the guest system and identify target process by page directory address. With the use of the QEMU plugin, we can get the code execution information of the target process, and detect and control the state of target process by several special network requests. These findings provide a theoretical foundation for implementing gray-box fuzzing specifically for network applications within firmware.

Based on the aforementioned methods, we develops AFLNetSpy, a gray-box fuzz testing framework specifically designed for network applications within firmware, by integrating the AFL-SPY plugin of QEMU into AFLNet. AFLNetSpy enables efficient gray-box fuzz testing for network applications within firmware, which successfully extends the gray-box fuzzing testing capabilities of AFLNet to the firmware application field and plays a significant role in prompting fuzz testing research in the firmware filed.

This thesis begins by introducing the research background, significance, and related work in the fields fuzz testing and firmware security etc. We then explains the methods of using the AFL-SPY plugin of QEMU to obtain code excution information of target process within the guest system and using special network requests to detect and control the state of the target process. Next, we present the whole architecture of AFLNetSpy and then highlight

key design points of each module. Finally, we design two experiments to analyze and validate AFLNetSpy from the perspective of stability and performance.

Key Words: Greybox Fuzzing; IoT Security; Network Application

目 录

摘 要	I
Abstract.....	II
第 1 章 绪论	1
1.1 研究背景	1
1.2 研究意义	3
1.3 国内外研究概况.....	5
1.4 本论文组织结构.....	8
第 2 章 相关工作	9
2.1 AFL	9
2.2 AFLNet	11
2.3 TriforceAFL	12
2.4 DECAF-QEMU	14
2.4 ISPRAS-QEMU.....	15
2.5 本章小结	15
第 3 章 系统设计要点.....	17
3.1 代码执行信息的获取.....	17
3.2 目标进程的状态探测和控制.....	17
3.2.1 目标进程的状态探测.....	17
3.2.1 目标进程的关键控制.....	18
第 4 章 系统详细设计.....	19
4.1 整体架构	19
4.2 QEMU-SPY	20
4.3 系统调用回调.....	24
第 5 章 实验分析	27
5.1 环境准备	27
5.2 稳定性分析	27
5.3 性能分析	29
结 论	31
参考文献	32
致 谢	34

第1章 绪论

1.1 研究背景

在当今信息技术高速发展的背景下，软件安全得到了广泛的关注。随着软件规模和复杂度的不断提高，依靠人工审查和手动测试来确保软件的正确性和安全性变得不再可行。因此，自动化软件测试技术尤其是模糊测试技术，应运而生，并在近年得到越来越多的关注。通过不断改进和利用模糊测试技术，人们在桌面应用如浏览器、网络服务、办公软件等程序中找到了大量有意义的错误和漏洞，证明了模糊测试技术的实用性。

模糊测试(fuzz testing)是一种自动化测试技术，它通过向程序输入随机或半随机的数据，观察程序的响应并检测异常行为。模糊测试的目标是发现软件中的潜在漏洞和错误，以改进软件的质量和安全性。模糊测试的一个重要优势是它可以发现那些由于输入组合庞大而难以穷尽的漏洞，从而使得测试更加全面和高效。

关于模糊测试的分类有两种指标，一是生成测试用例的方法，二是对待测程序信息的需求程度。

根据生成测试用例的方法，可将模糊测试划分为两类：①基于生成，基于生成的模糊测试是指模糊测试引擎根据用户规定的格式从头开始生成全新的测试用例。生成测试用例的过程可以基于随机算法、语法模型或其他规则。这种方法能够产生具有多样性和高覆盖性的测试用例，以探索待测程序的各个边界和异常情况。基于生成的模糊测试通常需要一定的领域知识和规范定义，以确保生成的测试用例符合预期的输入格式和语义；②基于变异，基于变异的模糊测试是指通过对已有测试用例进行变异来生成新的测试用例。用户首先提供一个或多个初始测试用例作为种子，然后通过对这些种子用例进行变异操作，如位翻转、删除、插入、替换等，生成新的测试用例。这种方法可以在已有测试用例的基础上进行迭代改进，更加高效地探索待测程序的执行路径和边界情况。基于变异的模糊测试通常需要一些变异操作的策略和机制来确保生成的测试用例具有多样性。

根据对待测程序信息的需求程度，可将模糊测试分为三类：①白盒模糊测试，白盒模糊测试需要使用待测程序的源码信息进行分析。测试人员或工具需要对待测程序的内部结构和逻辑有一定的了解，以便生成针对性的测试用例。白盒模糊测试能够

更深入地探索程序中隐藏的路径和边界条件，发现潜在的漏洞和异常情况。为了获取源码结构信息，通常需要进行静态分析、符号执行或动态插桩等技术；②灰盒模糊测试信息，灰盒模糊测试不需要对待测程序的源码进行分析，但需要获取和使用其他能够反映测试程序执行情况的信息。例如，可以利用程序的运行日志、覆盖率信息、异常处理机制等来指导测试用例的生成和选择。灰盒模糊测试可以在不了解程序内部细节的情况下，通过观察程序的行为和响应来推测潜在的漏洞和异常情况；③黑盒模糊测试：黑盒模糊测试不需要待测程序的任何信息。测试人员只需要了解待测程序的输入输出规范和预期行为，而无需关注程序的内部实现细节。黑盒模糊测试可以模拟真实用户的输入，发现程序在不确定或异常输入下的响应情况。黑盒模糊测试是最常见和通用的模糊测试方法，适用于各种类型的软件和系统。

与传统的手动测试和静态分析相比，模糊测试具有许多优势。首先，它可以自动化执行，减少了人工测试的工作量。其次，它可以探索更广泛的测试空间，因为输入是随机生成的，可以涵盖更多的边界情况和异常情况。此外，模糊测试还可以帮助发现未知的漏洞，甚至是零日漏洞，这些漏洞在软件发布之前都是未知的。

嵌入式系统和物联网设备的兴起与智能化的发展密切相关。嵌入式系统是指内置在其他设备或系统中的计算机系统，它们通常运行在嵌入式处理器上，用于控制和管理设备的各种功能。物联网设备是指通过互联网连接的各种智能设备，如智能家居、智能城市基础设施、智能工业设备等。这些设备在现代社会中起着越来越重要的作用，涵盖了各个领域，包括交通、医疗、能源、环境监测等。嵌入式系统和物联网设备中的固件程序是实现设备功能的核心软件组件。固件程序通常负责设备的控制、通信、数据处理等任务。由于嵌入式系统和物联网设备的特殊特点，固件程序的安全性至关重要。一个存在漏洞的固件程序可能会导致设备失效、数据泄露、远程攻击等严重后果。因此，对嵌入式系统和物联网设备中的固件程序进行充分的测试和验证是至关重要的。

然而，传统的模糊测试技术主要针对桌面应用程序进行设计 and 应用，对于嵌入式系统和物联网设备中的固件程序来说，存在一些特殊的挑战，这给模糊测试技术的应用带来了一定的困难：

(1) 嵌入式系统和物联网设备通常具有资源受限的特点。这些设备的处理能力和内存容量有限，无法承受大规模的模糊测试。传统的模糊测试技术通常需要大量的输入数据和计算资源来生成和执行测试用例，这在资源受限的环境下是不可行的。

(2) 固件程序通常在特定的硬件平台上运行，与桌面应用程序的执行环境有所不同。这意味着传统的模糊测试技术需要适应不同的硬件平台和操作系统，并针对特定的固件程序进行测试。这对测试工具和测试方法提出了更高的要求，需要充分考虑到嵌入式系统和物联网设备的特殊性。

(3) 嵌入式系统和物联网设备中的固件程序通常与外部环境和其他设备进行交互。这种交互可能涉及各种通信协议和接口，如无线通信、以太网、蓝牙等。传统的模糊测试技术需要能够模拟和生成各种通信协议和接口的测试数据，以覆盖不同的交互场景。这对测试工具和测试方法的设计提出了更高的要求。

针对以上挑战，研究人员和工程师们已经开始探索 and 开发适用于嵌入式系统和物联网设备的固件程序的模糊测试技术。一种常见的方法是通过测试用例生成和执行过程进行优化，以减少资源消耗。例如，可以利用针对嵌入式系统和物联网设备的特定特征和约束条件，设计高效的测试用例生成算法。同时，可以采用轻量级的测试工具和测试框架，以减少计算和存储资源的使用。另一种方法是针对特定硬件平台和操作系统进行测试工具和测试方法的设计和开发。这包括对硬件平台的深入了解，以便更好地理解固件程序的执行环境和特征。同时，还可以开发针对特定硬件平台和操作系统的模糊测试工具和测试框架，以提高测试效率和准确性。

此外，针对嵌入式系统和物联网设备中的通信协议和接口，也需要专门的测试技术和工具。这包括对通信协议和接口的解析和分析，以及针对特定协议和接口的测试用例生成和执行。可以利用模糊测试技术生成各种异常和边界情况的测试数据，以测试通信协议和接口的健壮性和安全性。另外，还可以考虑结合静态分析和模糊测试技术，以提高测试的效率和准确性。静态分析可以用于分析固件程序的代码和结构，发现潜在的漏洞和错误。结合模糊测试技术，可以针对性地生成测试用例，以验证静态分析的结果，并发现更多的漏洞和错误。

1.2 研究意义

随着物联网技术的快速发展和广泛应用，固件系统的安全性成为了一个日益重要的问题。固件是嵌入在各种物联网设备中的软件，负责控制设备的各种功能和操作。然而，由于固件系统通常存在漏洞和安全风险，黑客可以利用这些漏洞入侵设备，窃取用户信息，甚至控制设备进行恶意操作。因此，保护固件系统的安全对于确保物联网设备的可靠性和用户数据的隐私至关重要。

传统上，针对桌面系统程序的测试方法主要包括静态分析、动态分析和模糊测试。模糊测试是一种常用的自动化测试技术，通过向目标系统输入具有随机或半随机数据的输入，以尝试触发潜在的漏洞和错误。然而，由于固件系统的特殊性，传统的模糊测试技术和工具往往无法直接应用于固件系统。因此，将传统的模糊测试技术和工具扩展到固件系统，对于固件安全领域的发展具有重要的意义。

为了应对固件系统的特殊性，研究人员和安全专家们已经开始探索将传统的模糊测试技术和工具扩展到固件安全领域。一些针对固件系统的特定模糊测试工具和框架已经被开发出来，以支持固件系统的安全测试。这些工具和框架通常考虑了资源消耗、硬件架构和访问限制等因素，并提供了特定于固件系统的测试方法和技术。

本文提出了利用QEMU的全系统仿真模式和插件功能，收集代码执行信息和控制测试程序状态的方法。QEMU是一个开源的模拟器，可以模拟多种计算机体系结构的硬件环境。通过在QEMU中实现插件功能，可以监控和收集固件系统运行过程中的代码执行信息，从而为后续的模糊测试提供基础数据。此外，该方法还可以控制测试程序的状态，以便更好地触发潜在的漏洞和错误。基于这一方法，本文开发了原型系统AFLNetSpy，该系统支持对固件网络应用进行灰盒模糊测试。

AFLNetSpy是在已有的灰盒模糊测试工具AFLNet的基础上进行扩展而来的。AFLNet是一个专门用于网络应用和网络协议的灰盒模糊测试工具，通过模拟网络请求和响应，在初始测试用例的接触上通过随机变异生成具有新的测试用例，并能够监控应用程序的行为。通过将AFLNet扩展到固件网络应用领域，本文为固件系统的安全性提供了一种全新的测试方法。AFLNetSpy的开发不仅充分利用了QEMU的全系统仿真模式和插件功能，还针对固件系统的特殊性进行了相应的优化和改进。

这一研究成果具有重要的意义和广泛的应用前景。首先，AFLNetSpy为固件安全领域提供了一种新的测试工具和方法。传统的模糊测试技术和工具往往无法直接应用于固件系统，而AFLNetSpy通过扩展现有的工具，充分考虑了固件系统的特殊性，为固件安全领域的研究和实践提供了一种创新的解决方案。其次，AFLNetSpy的应用可以提高固件系统的安全性和可靠性。通过对固件网络应用进行灰盒模糊测试，可以发现潜在的漏洞和安全风险，并及时采取相应的措施进行修复和改进，从而提高固件系统的抗攻击能力和稳定性。

此外，AFLNetSpy的研究成果还为后续固件安全领域的研究提供了可供借鉴的思路和方法。随着物联网技术的不断发展，固件安全的重要性也日益凸显。研究人员

和安全专家需要不断探索和创新，以应对新的威胁和挑战。AFLNetSpy提出的基于QEMU的全系统仿真模式和插件功能的方法，为固件安全领域的研究提供了一种新的思路和技术手段。其他研究人员可以在此基础上进行进一步的研究和改进，推动固件安全技术的进一步发展。

最后，AFLNetSpy的研究成果对于推动物联网技术和固件安全技术的进一步发展具有积极的意义。随着物联网设备的普及和应用场景的不断扩大，固件安全的重要性将变得越来越突出。通过提供一种有效的测试工具和方法，如AFLNetSpy，可以帮助设备制造商和开发者提高固件系统的安全性和可靠性，从而促进物联网技术的健康发展。同时，通过推动固件安全技术的研究和创新，可以有效应对不断增长的网络安全威胁，保护用户的隐私和数据安全，推动物联网技术的广泛应用和进一步发展。

总的来说，本文提出的基于QEMU的全系统仿真模式和插件功能的方法以及开发的原型系统AFLNetSpy在固件安全领域具有重要的研究意义和实际应用价值。它不仅为固件安全领域提供了新的测试工具和方法，有助于提高固件系统的安全性和可靠性，而且为后续该领域的研究提供了可供借鉴的思路，对于推动物联网技术和固件安全技术的进一步发展具有积极意义。

1.3 国内外研究概况

1988年，Miller首次提出模糊测试的基本概念，即通过反复向程序提供随机输入数据来寻找崩溃^[1]。之后的三十多年里，模糊测试技术不断发展，历经多个发展阶段，现已成为成为一种广泛使用的自动化测试和漏洞挖掘技术。

模糊测试发展早期，相关研究主要集中在黑盒测试。在2013年AFL^[2]诞生之后，基于覆盖率引导的灰盒模糊测试研究成为主流。AFL是由Google团队开发的一种基于变异的灰盒模糊测试工具，它通过插桩获取测试目标的代码执行信息，并在此基础上实现覆盖率驱动的用例选择和变异算法。本文将在2.1节详细介绍AFL的工作流程。

基于AFL，国内外研究者进行了大量的改进研究。其中，Marcel Böhme在AFLFast^[3]中将基于覆盖率的灰盒模糊测试建模为对马尔科夫链状态空间的系统探索，并提出了几种加强对低频路径重视程度的搜索和调度算法，提高了灰盒模糊测试的效率。Chenyang Lyu等人在MOpt^[4]中利用定制的粒子群优化(PSO)算法，找到算子在模糊测试有效性方面的最优选项的概率分布，并提供“起搏器”模糊测试模式，以加速PSO的

收敛速度，实现模糊测试效率的提高。**LAF-INTEL**^[5]提出将多字节数据比较转化为逐字节数据比较的方法，以解决模糊测试中的大数和字符串比较问题。**RedQueen**^[6]基于**KAFL**^[7]提出了绕过模糊测试中硬比较与校验和检查的方法。

AFL存在的一个主要限制是对输入数据的结构缺乏感知能力，只能够实现比特级的变异操作。**Peach**^[8]是一个能够感知测试数据结构的黑盒模糊测试工具，它能够通过用户规定的输入模型拆解有效的测试数据，并将得到的片段进行删除、修改和重新组合以得到新的测试数据。**Pham**等人在**AFLSmart**^[9]中集成了**Peach**的输入结构组件和**AFL**的覆盖反馈组件，可实现对测试用例高度结构化的变异，并保证新测试用例的有效性，适用于遵循树结构的所有复杂数据格式。

AFL已于2021年停止维护，大量基于**AFL**的学术研究和改进工作的成果处于分散状态。为解决这个问题，**Andrea Fioraldi**等人提出**AFL++**^[10]，它是一个社区驱动的开源工具，整合了大量基于**AFL**和最新的模糊测试研究成果。同时通过提供可自定义的变异API和插件机制，**AFL++**为未来的模糊测试研究提供了良好的基础设施，使得研究者能够快速开发原型系统以及尝试组合各种改进策略，并进行统一公平的评估，对促进模糊测试领域的健康发展具有重要意义。

尽管**AFL++**提供的自定义API和插件机制为后续的模糊测试研究提供了基础设施，但由于受制于**AFL**的基础代码框架，难以在软件工程水平上保证**AFL++**的扩展性。随着越来越多的改进算法加入到其中，**AFL++**项目的代码量逐渐增大，变得越来越难以维护，且延续了多个新分支间相互隔离的历史问题。

为此，**AFL++**的开发团队**Andrea Fioraldi**等人在**AFL++**之外，以可扩展性和可复用性为第一原则，使用**RUST**从头实现了**LibAFL**^[11]项目。**LibAFL**由一组库组成，用户可以通过组合基于可扩展实体的库组件构建自定义模糊器。它实现这一目标要归功于几个因素：①易于扩展，**LibAFL**的设计目标之一是提供一个易于扩展的模糊测试框架。为了实现这一点，**LibAFL**利用了**Rust**编程语言的特性，如所有权系统、模式匹配和模块化设计。这些特性使得开发人员能够轻松添加新的功能、算法和组件，并能够快速构建和测试新的组件。此外，**Rust**的强类型系统和内存安全性保证也有助于减少错误和漏洞的可能性，提高代码的可靠性和稳定性；②基于对现代模糊器组件的分类，**LibAFL**提供了一组库组件，包括输入生成器、覆盖率分析器、变异器等。用户可以根据其需求选择和组合这些组件，构建出适用于特定应用场景的定制化模糊器。这种组件化的设计使得**LibAFL**更加灵活和可定制，可以根据用户的具体需

求进行定制化的模糊测试；③充分利用RUST语言特性，LibAFL利用了Rust在编译时简单快速地序列化对象和组件插槽的功能，这种序列化机制使得LibAFL能够高效地处理和传递复杂的数据结构，提高了模糊测试的性能和效率。同时，这种序列化的方式也有助于减少运行时的开销，进一步提高了模糊测试的速度和响应能力；④开源驱动：作为一个开源项目，LibAFL积极关注研究和创新，并不断更新和改进其算法和功能。它与模糊测试研究社区保持紧密联系，吸纳最新的模糊测试技术和最佳实践。通过提供最新的算法和功能，LibAFL能够帮助用户在模糊测试过程中获得更好的测试覆盖和漏洞发现能力，对模糊测试领域的现代化发展以及打通学术界和工业界的壁垒具有重要意义。

除了提高模糊测试流程各个环节的效率，将模糊测试扩展到更多的领域也是主要研究方向之一。

网络应用和网络协议的安全性一直是互联网领域的关注焦点。由于网络应用和协议的复杂性和多样性，传统的测试方法往往无法覆盖所有潜在的漏洞和安全隐患。因此，灰盒模糊测试成为一种被广泛采用的方法，它通过模拟输入数据的多样性和异常性，尝试触发目标应用和协议中的漏洞和错误行为。在网络应用和网络协议领域，Thuan Pham等人率先提出AFLNet^[12]，将AFL的灰盒模糊测试能力扩展到了网络应用和网络协议的测试中，本文将在2.2节详细介绍AFLNet的工作流程。Anastasios Andronidis等人在AFLNet的基础上提出改进策略，并实现SnapFuzz^[13]原型系统，加快了网络应用模糊测试的效率。

在固件模糊测试领域，NccGroup率先提出TriforceAFL^[14]，将AFL的灰盒模糊测试能力成功应用于固件系统的系统调用测试中。TriforceAFL的关键创新在于结合了QEMU的全系统仿真能力。QEMU^[15]是一种开源的虚拟化软件，可以模拟多种硬件平台和操作系统环境。通过与QEMU的结合，TriforceAFL可以在模糊测试过程中对固件系统进行全面的仿真和监控，以便更好地模拟真实设备环境和检测系统调用的执行情况。本文将在2.3节详细介绍TriforceAFL的工作流程。Yaowen Zheng等人在FirmAFL^[16]和EQUAFL^[17]中分别提出利用混合仿真模式和增强的QEMU用户仿真模式两种方法，以提高固件应用程序模糊测试的效率，然而这些方法对目标程序有一定的限制，存在不够普适的问题。

此外，还有一些基于机器学习和人工智能的模糊测试方法正在被研究和开发。这些方法利用机器学习算法实现测试用例进行生成、变异和选择，以提高模糊测试的效

率和覆盖率。例如，使用遗传算法、强化学习等技术来指导测试用例的生成和变异。这些方法在某些情况下可以产生更好的测试用例，但也面临着挑战，如测试用例的可解释性、训练数据的获取等问题。Ruijie Meng等人在ChatAFL^[18]工作中，意识到大语言模型在训练过程中很可能已经提取了数百万页的人类可读的协议规范，可以在协议模糊测试期间利用大语言模型关于网络协议的知识，于是他们开发了一个由大模型引导的协议实现引擎供模糊测试使用，ChatAFL能够为目标协议中的没用中消息类型构建相应的语法，然后通过与大语言模型的交互来改变消息或预测消息序列中所需的下一条消息。

总的来说，目前模糊测试领域，尤其是针对固件应用的模糊测试领域，发展十分迅速，但仍存在一些问题供后续研究者探索解决。

1.4 本论文组织结构

本论文的正文部分按照以下六个章节组织划分：

(1) 第一章：绪论。重点介绍本论文的研究背景和研究意义，并阐明了模糊测试和固件模糊测试领域的国内外研究现状。同时介绍了本论文提出的新方法和开发出的AFLNetSpy原型系统。

(2) 第二章：相关工作。本章重点介绍模糊测试领域的知名工具AFL，以及基于AFL进行扩展得到的AFLNet和TriforceAFL。同时介绍DECAF和ISPRAS-QEMU两个利用QEMU插件系统实现虚拟机内省的相关工作。

(3) 第三章：系统设计思路。本章重点介绍了本文提出的两个新方法：①利用QEMU插件系统获取进行灰盒模糊测试所需的目标程序代码执行信息；②利用网络请求探测和控制目标程序状态。同时介绍了系统设计相关的其它注意事项。

(4) 第四章：系统详细设计。本章首先介绍了AFLNetSpy系统的整体架构，然后对关键模块QEMU-SPY的内部实现进行讲解。

(5) 第五章：实验分析。本章首先介绍实验环境的准备工作，然后从稳定性和性能两个方面设计实验对AFLNetSpy原型系统进行分析。

(6) 第六章：结论。本章根据前文的方法和实验结果，对研究成果进行总结，并指出不足之处，同时对未来进行展望。

第2章 相关工作

本章选择了几项和本文研究内容最相关的工作，包括模糊测试领域的里程碑AFL、网络应用模糊测试领域的先驱性工作AFLNet、固件模糊测试领域的先驱性工作TriforceAFL，以及基于QEMU插件系统实现虚拟机内省的两项工作DECAF-QEMU和ISPRAS-QEMU，进行详细介绍。

2.1 AFL

AFL是Google团队开发的一个知名的灰盒模糊测试工具，通过收集测试进程的代码执行信息，并利用遗传算法，进行覆盖率引导的灰盒模糊测试。AFL支持两种使用模式，静态插桩模式和动态插桩模式。

静态插桩模式适用于有待测程序源码的情况。AFL对gcc等编译工具进行包装，得到afl-gcc等工具。使用AFL包装后的编译工具对待测程序进行编译时，AFL会在生成的程序中添加一些代码片段，用于在后续测试过程中收集代码执行信息，以及探测和控制程序运行状态，并在模糊测试过程中和afl-fuzz保持交互。静态插桩模式下，AFL的工作流程如下：

- (1) 首先使用afl-fuzz命令启动模糊测试后，afl-fuzz首先会创建一个子进程并执行待测程序，待测程序启动后作为Forkserver和afl-fuzz保持通信。
- (2) 之后的每一次测试前，afl-fuzz会通过管道请求Forkserver创建一个测试进程，并等待测试进程的进程号。
- (3) Forkserver接收到afl-fuzz的请求后，会创建测试进程，并将其进程号通过管道返回给afl-fuzz，实现同步。
- (4) afl-fuzz接收到测试进程号后，从测试队列选择或变异生成一个测试用例，并通过系统初始化阶段设置好的文件描述符将测试用例发送给测试进程。然后等到Forkserver返回测试进程的退出状态码，若定时器超时，则主动kill掉测试进程。测试进程执行过程中会不断更新trace_bits数组，以记录代码执行情况。
- (5) Forkserver会等待其子进程即测试进程退出，并将退出状态码返回给afl-fuzz进行分析，实现同步。
- (6) afl-fuzz在接收到测试进程的退出状态码后，会检查通过共享内存和测试进程共享的trace_bits数组的状态，并根据trace_bits的状态和测试进程退出状态码对本次测试用

例的价值进行评判，如果触发了新的执行路径，则将该测试用例保存至测试队列。然后回到第(2)步，重复(2)-(6)的过程，直到用户手动停止afl-fuzz。

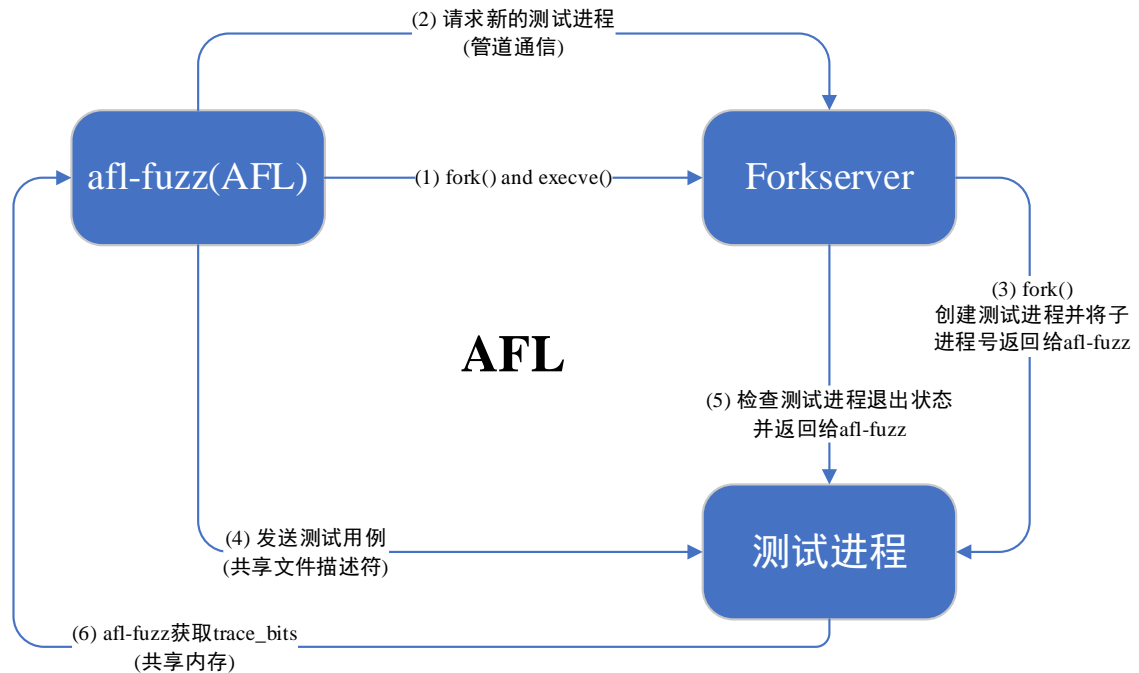


图 2-1 AFL 工作流程图

动态插桩模式适用于无法访问待测程序源码，只有二进制程序的情况。**AFL**在动态插桩模式下的工作流程和静态插桩模式基本一致，但是这种情况下由于无法通过在待测程序中插入二进制片段，没办法探测和控制测试进程的状态，也无法测试进程的代码执行信息。

为了解决这个问题，**AFL**引入并修改了用户模式的**QEMU**。修改后的**QEMU**可以发挥和静态插桩模式下**Forkserver**一样的作用，能够和**afl-fuzz**通过管道保持通信和状态同步。同时由于**QEMU**的翻译执行机制中翻译块的存在，代码执行信息的收集也变得可行。

另外需要注意的是，动态插桩模式的效率要比静态插桩模式要低，且动态插桩模式无法像静态插桩模式那样，根据跳转指令切分程序代码块，所得到的代码执行信息的意义不如静态插桩模式直观。但**QEMU**的使用使得测试不同架构下的二进制程序变得可行，扩大了**AFL**的适用范围。

2.2 AFLNet

AFLNet是Thuan Pham等人针对网络应用测试开发出的AFL扩展版。AFL只支持测试以文件或命令行的方式进行输入的程序，如果想要使用AFL测试某个网络应用，首先必须修改该应用的输入方式，使其能够从文件中读取网络请求。当没有源代码时，AFL就对该网络应用束手无策了。而AFLNet在AFL的基础上添加了网络通信支持，能够通过网络套接字直接向待测网络应用发送测试请求。同时，考虑到网络应用的错误往往需要一个序列的请求才能够触发，因此AFLNet不像AFL一次测试只发送一个请求，它支持在一个测试用例中存放多个请求报文，并在一次测试中按顺序发送它们。

AFLNet的工作流程的如图所示，和AFL的工作流程基本一致。区别主要有两点：

(1) 一是第(4)步中AFLNet采用网络套接字即socket通信的方式向测试进程发送测试请求，然后接受响应，并提取响应中的状态码，状态码不仅可以用于协议状态机的分析，还能够和代码执行信息一起用来评估测试用例的价值和指导新测试用例的生成。

(2) 二是由于网络应用通常处于无限循环状态，处理完一次测试用例后并不会自动退出，为此AFLNet支持两种策略：①一个测试进程只处理一个测试用例，在一次测试结束后，由afl-fuzz主动向测试进程发送SIGTERM信号以关闭测试进程；②允许一个测试进程处理多个测试用例，直到出现超时或者崩溃，再启动一个新的测试进程。本文开发的原型系统AFLNetSpy就采用了第二种策略。

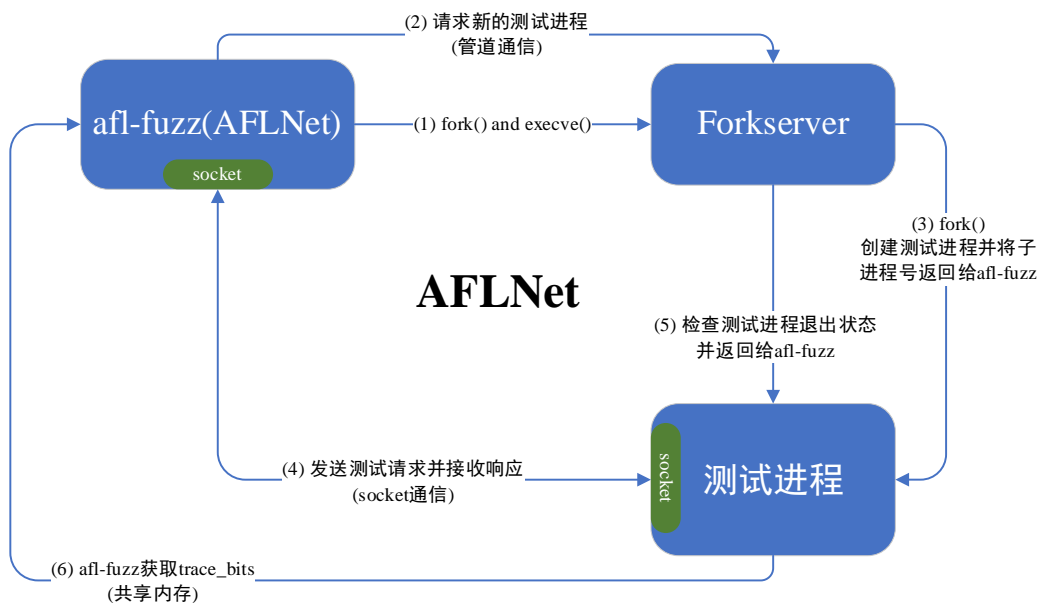


图 2-2 AFLNet 工作流程图

同时，AFLNet提出一个新的协议推断算法，通过观察AFLNet和待测网络应用间交换的请求响应数据，能够一定程度上还原出通信采用的协议格式。因此，作为一个协议测试工具，它并不需要用户手动指定协议规范和报文格式，可以直接将用户收集到的客户端和待测服务器的通信流量中的请求作为种子，在重放请求的过程中，自动学习协议知识。不过，当初始测试用例较少时，自动获取的报文格式难免会存在误差，导致变异生成的大部分测试用例仍然是无效的。

另外需要注意的是，AFLNet在发送一次测试用例后，会尝试接收响应。但是，以HTTP服务器为例，当测试请求无效时，服务器会直接丢弃该请求报文，并不会返回响应。因此，相比于AFL，AFLNet需要用户额外手动设置一个合适的等待响应的最大时间。

2.3 TriforceAFL

TriforceAFL是NccGroup针对固件系统调用测试开发出的AFL扩展版。它创新性地成功将QEMU的系统模式和AFL结合起来，使得TriforceAFL能够对QEMU模拟启动的固件系统中的系统调用进行高效的灰盒模糊测试。

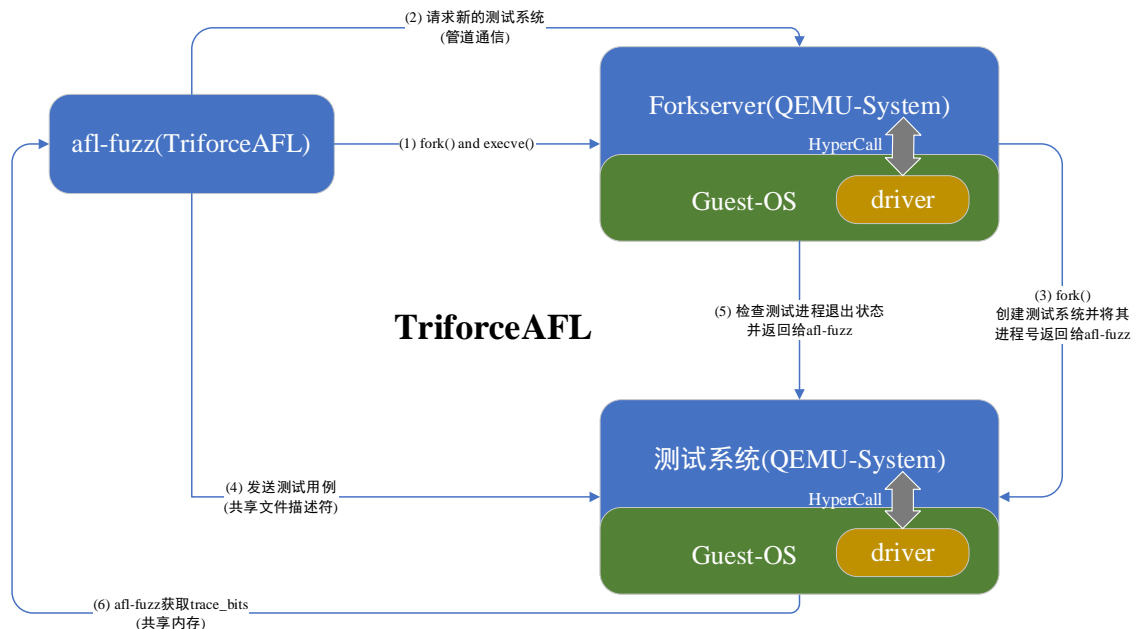


图 1-3 TriforceAFL 工作流程图

TriforceAFL的整体工作流程如图所示，和AFL工作流程的主要区别，也即TriforceAFL重点解决的问题有三个：

(1) 一是测试用例的输入。在使用AFL对桌面应用程序进行的模糊测试中，能够通过dup()函数在父子进程之间，主要是afl-fuzz进程和测试进程之间，共享所需的文件描述符，并通过该文件描述符直接传递测试用例。然而运行在主机系统上的afl-fuzz，无法与运行在QEMU内部的客户机系统及其中的应用程序共享文件描述符，也就无法通过这种方式来传递测试用例。出于进行固件系统调用测试的目的，TriforceAFL采用的策略是，在固件系统中添加一个用户编写的driver程序，driver程序能够通过执行特殊的HyperCall指令与QEMU进行协同，而QEMU进程作为afl-fuzz的子进程，能够通过管道和afl-fuzz进程保持通信并共享文件描述符。于是，以QEMU为桥梁，客户机系统中运行的driver程序便能够接收到afl-fuzz发送的测试用例。

(2) 二是获取测试用例执行结果并退出测试系统。driver在接收到测试用例后，能够启动一个子进程作为测试进程执行测试用例对应的系统调用，并获取其执行状态，然后通过特定的HyperCall指令将执行状态以QEMU作为媒介，返回给afl-fuzz进程。同时，driver还会通过另一条特定的HyperCall指令，通知当前用于测试的QEMU进程退出，QEMU进程退出后，其内部用于本次测试的客户机系统自然也就实现了退出。

(3) 三是测试系统的状态恢复。由于系统调用测试的特殊性，异常的系统调用很可能导致客户机操作系统的崩溃。为保证模糊测试的持续运行，每一次测试结束后，必须重启客户机系统或设法恢复客户机系统状态至执行测试用例前。重启客户机系统耗时过长，显然不是一个可行的方案。TriforceAFL通过实现QEMU-System级别的Forkserver，实现了客户机系统状态的快照和恢复。在客户机系统启动完成并运行driver进程后，driver会通过一个特定HyperCall指令告知QEMU-System客户机系统已经就绪，然后QEMU-System就会保存当前的虚拟CPU状态，后续创建的每一个测试系统都将以此刻为运行起点。

需要注意的是，虽然NccGroup宣称依靠TriforceAFL可以实现“Fuzz Everything”，TriforceAFL也的确提供了创新性的思路，但本文经过测试，发现事实并非如此。目前依靠TriforceAFL仅仅只能够实现固件系统调用级别的灰盒模糊测试，要想实现固件内应用程序级别的灰盒模糊测试，还需要额外的工作。原因如下：

(1) 代码执行信息的收集：QEMU系统模式下，启动的客户机系统中往往同时运行着大量的应用程序，而QEMU只会根据当前虚拟CPU的状态来获取二进制代码块，并进

行翻译和执行，无法直接判断出当前代码块属于客户机系统中的哪个进程。为此，TriforceAFL采用的策略是，通过修改固件的文件系统镜像和初始化脚本，保证客户机系统启动后其内部只有driver进程和其子进程即测试进程处在运行状态，然后通过统计高地址空间即操作系统使用的地址空间中的指令执行情况，获取测试用例对应的代码执行信息。显然，这种方式并不适用于固件内的普通应用程序。

(2) QEMU级别的Forkserver: TriforceAFL通过QEMU级别的Forkserver来保证每一次测试结束后，下一次测试执行前，能够恢复客户机系统的状态。然而，经过测试，这种方式并不能够在恢复客户机系统状态的同时恢复客户机内各个应用程序的执行状态。本文以HTTP服务器为例进行了测试，当客户机内的HTTP服务器接收到网络请求，即在QEMU层监测到该服务器进程调用accept系统调用时，仿照TriforceAFL的实现方式，退出vcpu线程，在主线程中执行fork()命令，得到的子进程无法继续处理刚才接收到的网络请求，且有可能触发各类段错误。因此，对于固件内应用程序级别的模糊测试，采用QEMU级别的Forkserver并不是可行的方案。本文将提供一种针对固件内网络应用的替代方案。

2.4 DECAF-QEMU

DECAF^[19]是Andrew Henderson等人开发的一个动态可执行代码分析框架。DECAF为QEMU提供了一套易于使用的、事件驱动的插件机制，并基于该机制实现了实时的虚拟机内省即VMI功能。

动态二进制分析是程序分析中研究中广泛采用且不可或缺的技术。在DECAF之前，尽管已经提出了几种动态二进制分析工具和框架，但它们都存在一些问题，包括性能下降严重、分析代码与被分析程序之间存在语义差距、特定于特定体系结构/操作系统、仅支持用户模式、缺乏API等。为此，Andrew Henderson等人在QEMU的基础上，构建了DECAF，实现基于虚拟机的多目标、全系统动态二进制分析框架，并提供即时虚拟机内省(Just-In-Time Virtual Machine Introspection)能力，支持基于插件的简单易用的事件驱动编程接口。通过结合一种新颖的TCG指令级别的位粒度污染分析技术，DECAF能够实现对TCG指令的精细控制，并在此基础上，提供了三个与平台无关的插件，分别用于监测特定指令、监测API执行情况和监测键盘输入情况。尽管DECAF的主要目的是实现比特粒度、高准确性的TCG指令污点分析，但具备低开销VMI功能的DECAF-QEMU也可作为独立组件用于其它目的。

另外需要注意的是，DECAF-QEMU所使用的QEMU版本较低，难以启动较新的固件镜像如OpenBMC等镜像，且对操作系统的内核结构存在依赖，当特定数据结构出现变动时需要重新进行修改适配。但其实现方法仍然具有很大的参考价值，本文开发出的原型系统AFLNetSpy所需的系统调用监测功能就借鉴了其实现思路。

2.4 ISPRAS-QEMU

ISPRAS-QEMU^[20]是Ivan Vasiliev等人受ISPRAS组织支持开发出的一个基于QEMU的非入侵式虚拟机插桩和内省框架。它是一个用于对虚拟机内二进制程序进行动态分析的轻量级框架，具备跨系统和非入侵即不修改目标二进制程序及其执行流程两个主要特点。

ISPRAS-QEMU通过建立分层的事件驱动的插件机制，使得它实现的整个VMI系统结构清晰且易于扩展。然而和DECAF-QEMU一样，每个插件都需要在程序启动后手动开启，而不能直接在命令行参数中进行设置，在某些场景下使用起来不够方便。

另外，和DECAF-QEMU相比，ISPRAS-QEMU不需要依赖操作系统内核结构的信息，而是通过ABI规范创建了一些可以恢复内核级别信息的工具，从而实现各个事件的监测和定位。这种策略使得ISPRAS-QEMU具备了不依赖系统板的特性，但同时也损失了一些功能和便利性，如只能通过页目录地址标识进程，以至于无法通过指定进程名称和进程号的方式来筛选属于特定进程的各类事件。

2.5 本章小结

本章介绍了模糊测试和利用QEMU构建VMI系统的相关工作。其中，模糊测试方面，介绍了AFL、AFLNet、TriforceAFL三者的工作流程、各自特点及注意事项。利用QEMU构建VMI系统方面，介绍了DECAF-QEMU和ISPRAS-QEMU两项通过构建QEMU插件系统实现VMI功能的工作及其各自特点和注意事项。

本文将基于上述相关工作，进行一定的改进和扩展。上述相关工作中的AFLNet和TriforceAFL等工具对于固件内普通应用程序的灰盒模糊测试束手无策，而本文基于AFLNet和修改后的QEMU开发出的AFLSpyNet系统能够实现固件内网络应用的灰盒模糊测试。第三章和第四章将分别详细介绍AFLNetSpy的设计要点和模块细节。

另外相较于DECAF-QEMU和ISPRAS-QEMU实现的VMI内省系统，本文的主要改进点包括：

- (1) 基于最新稳定版的QEMU进行开发，保证系统能够支持较新的固件镜像如OpenBMC等镜像的启动。
- (2) 基于QEMU官方提供的插件机制进行开发。QEMU官方从4.2版本起开始提供一套插件机制，目前已经比较稳定，但功能较少，需要进行二次开发。使用QEMU官方的插件机制，而不是像DECAF-QEMU或ISPRAS-QEMU一样自行实现或使用第三方提供的插件系统，可以保证系统稳定性和对QEMU版本的向后兼容性。

第3章 系统设计要点

本章将讲解AFLNetSpy系统的设计要点：①目标进程代码执行信息的获取；②目标进程的状态探测和控制。

3.1 代码执行信息的获取

要想进行固件内应用程序的灰盒模糊测试，首先需要能够获取目标进程的代码执行信息。

要实现这个任务，第一步是设法区分不同进程的指令。Linux系统下，每启动一个进程，会为其分配一个页目录，页目录地址和进程是一一对应的对应关系，因此可以通过页目录地址来标识不同的进程。且当前正在执行的进程的页目录地址会被存放在VCPU的特定页目录地址寄存器内，根据翻译块执行时VCPU页目录地址寄存器的值，即可判断当前翻译块中的指令属于哪一个进程。

第二步是要设法获取目标进程对应的页目录地址。考虑到目标进程为网络应用，通过向客户及系统内的目标进程发送一个网络请求，然后在QEMU中监测接下来触发Accept系统调用时VCPU的页目录地址寄存器，即可获得目标进程对应的页目录地址。之后在QEMU中就能够根据当前执行的翻译块对应的页目录地址是否和目标进程的页目录地址相同，来判断当前翻译块中的指令是否属于目标进程，从而筛选和获取目标进程的代码执行信息了。

3.2 目标进程的状态探测和控制

要想进行固件内应用程序的灰盒模糊测试，还要实现目标进程的状态探测和控制。

3.2.1 目标进程的状态探测

目标进程的状态探测，主要包括两个方面：

- (1) 判断客户机系统和目标进程是否启动完毕：通过监测execve等用于创建新进程的系统调用及其参数，即可根据进程名称判断目标进程是否启动。同时，可以通过进程名称选定一个大概率最后启动的进程，当该进程启动时，认为客户机启动完成。
- (2) 每次测试结束后检查目标进程是否仍处于运行状态：由于目标进程是网络服务器，除了检查它返回给afl-fuzz的响应中的状态码，还应在每次测试结束后检查目标进程

的存活状态，以保证当目标进程崩溃时及时得到通知。本文采取的策略是通过发送由用户编写的针对目标进程的TEST_ALIVE网络请求，根据返回情况判断目标进程是否仍处于运行状态。

3.2.1 目标进程的关键控制

目标进程的控制，关键点是目标进程的退出和重新启动。

对于目标进程的退出，经过测试，当外部请求触发客户机内目标进程的一个段错误时，会触发异常信号，目标进程会等待客户机系统处理该信号，并关闭目标进程。然而，由于QEMU的实现机制，异常信号不会立刻得到处理，目标进程也不会迅速被关闭，从触发段错误到目标进程成功退出，需要4分钟左右。这是难以接受的。

因此，本文提出一种主动探测式的方法，通过TEST_ALIVE请求探测目标进程状态，若为成功返回，则认为目标进程出现了CRASH，并向Spy-Agent程序发送强制杀掉目标进程的命令，从而实现目标进程的快速退出。其中，Spy-Agent是本文实现的一个需要内置于客户机系统内部运行，负责接收外部请求实现远程命令执行的网络程序。其实现细节会在第四章进行详细介绍。

对于目标进程的重新启动，为保证模糊测试的持续运行，在触发目标进程的崩溃并退出目标进程后，需要重新启动新的目标进程来继续模糊测试。这可以通过向客户及系统内的代理程序Spy-Agent发送重启目标进程的shell命令来实现，如systemctl restart \$TARGET_SERVICE等。

另外，还需要注意的是，现代操作系统及服务管理程序如systemd等，可能存在避免服务器频繁异常重启的保护机制，因此可能需要在强制退出目标进程后，重启目标进程前，重置该网络服务的错误状态信息，如使用systemd的systemctl reset-failed命令等。

第4章 系统详细设计

4.1 整体架构

AFLNetSpy的整体架构如图所示。其中trace_enabled和next_step与trace_bits一样，是afl-fuzz和QEMU-SPY通过共享内存方式共享的两个变量，用于控制和同步trace_bits的记录。QEMU-Spy为经过扩展的QEMU和AFL-SPY插件组成的具有VMI功能的系统。SpyAgent进程为本文实现的能够进行远程命令执行的网络服务进程，其具体实现可参见本文的代码仓库。

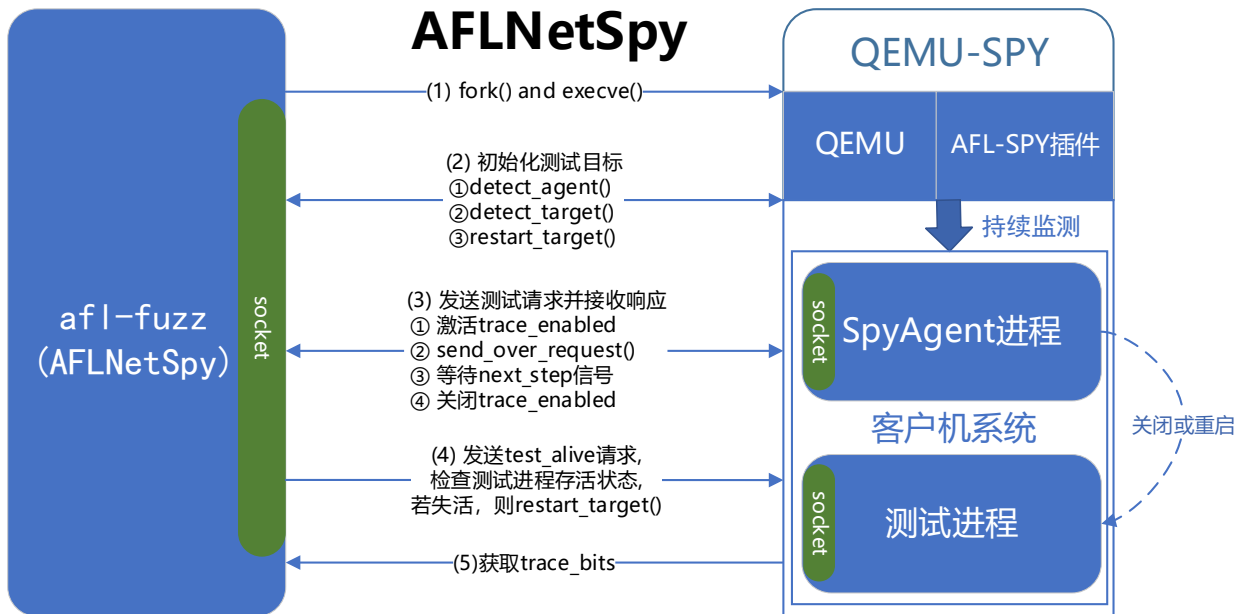


图 4-1 AFLNetSpy 系统架构图

结合上图，AFLNetSpy的整体工作流程如下：

(1) 首先afl-fuzz进程创建一个子进程用来运行QEMU，QEMU启动客户机系统前会加载通过命令行参数指定的AFL-SPY插件libaflspy.so，并注册所需的各类回调函数。启动后的客户及系统内，包含待测试的目标网络进程，和提前内置或后续传输进客户机系统的SpyAgent进程，该进程能够接收外部的网络请求并执行请参数中的shell命令，可以实现快速关闭和重启目标进程。

(2) 初始化测试目标，包含三个步骤：①detect_agent，向SpyAgent进程发送用户编写的TEST_AGENT请求，AFL-SPY插件在监测到该请求时，会将SpyAgent进程的页目录地址设为AgentCtx，并将该值通过管道返回给afl-fuzz进程；②detect_target，向目标测试进程发送用户编写的TEST_ALIVE请求，AFL-SPY插件在监测到该请求时，会将目标进程的页目录地址设为TargetCtx，并将该值通过管道返回给afl-fuzz进程；③restart_target，由于重启后的目标进程的页目录地址会发生改变，需要重新确定。因此该函数分为两步，一是向SpyAgent进程发送用户编写的RESTART_TARGET请求，SpyAgent接收到该请求后，会提取请求参数中的shell命令并执行，从而实现目标进程的重启，同时AFL-SPY插件在监测到该请求时会把TargetCtx的值设为0，表明原TargetCtx已失效，需要重新定位。二是执行②中提到的detect_target函数，从此实现TargetCtx的更新。

(3) 发送测试请求并接收响应，包含四个步骤：①激活trace_enabled，通知AFL-SPY插件，对接下来的目标进程的翻译块进行记录；②send_over_request是AFLNet提供的发送测试请求并接受响应的函数，本文对其进行了扩展，以支持HTTPS通信；③等待next_step信号，AFL-SPY插件在监测到目标进程执行特定系统调用如sendto、sendmsg时，将next_step的值设为1，表示此次测试请求已经处理完毕可以开始准备下一次测试了。afl-fuzz进程会轮询next_step的值，直到该值为1或达到最大轮询次数，则进入下一步；④关闭trace_enable，通知AFL-SPY停止记录翻译块的执行信息，保证得到的trace_bits的数据足够干净。

(4) 发送TEST_ALIVE请求，检查处理上次测试请求后目标测试进程的存活状态，若未正常返回，则认为上次的测试请求触发了测试进程的CRASH，然后使用restart_target函数重启一个新的测试进程，保证测试能够继续进行。若正常返回，说明上次的测试请求没能触发测试进程的CRASH，出于性能考虑，可以不重启测试进程。

(5) afl-fuzz获取AFL-SPY插件更新过的trace_bits，进行后续分析。然后生成新的测试请求，重复(3)-(5)步，直到用户手动停止测试。

4.2 QEMU-SPY

本节介绍QEMU-SPY系统的详细设计。QEMU-SPY系统由QEMU和AFL-SPY插件组成。

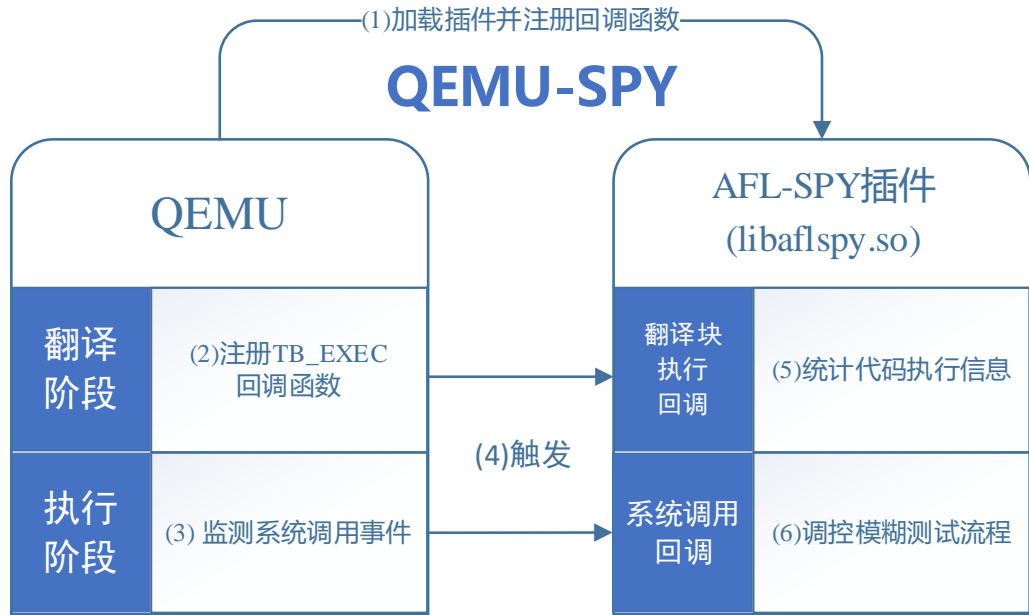


图 4-2 QEMU-SPY 系统结构图

QEMU-SPY的架构的如图所示，其工作机制如下：

- (1) 加载插件并注册回调函数：首先QEMU进程在启动后，会根据命令行参数加载指定的AFL-SPY插件，即libaflspy.so，并注册插件中的各个回调函数供后续执行。
- (2) 注册TB_EXEC回调函数：当客户机架构和宿主机架构部不同时，如客户机为arm架构，宿主机为x64架构，QEMU需要将客户机二进制代码翻译为宿主机二进制代码才能够执行。翻译阶段包含两个步骤：①将客户机二进制指令块翻译为TCG中间指令块；②将TCG中间指令块编译为宿主机二进制指令块。通过在每个TCG中间指令块中插入特定的代码，能够实现类似回调函数的功能，且QEMU提供的HELPER机制使得这项操作变得易于实现。本文通过在每个翻译块，具体来说是在TCG中间指令块中，使用QEMU提供的HELPER机制插入HELPER(tb_exec_spy)函数，使得每个翻译块执行前都会调用AFL-SPY插件中编写的翻译块执行回调函数，从而实现代码执行信息的统计。
- (3) 监测系统调用事件：本文还对QEMU进行了另一项扩展，使其支持系统模式下客户机调用的监测。具体实现方法如下图所示，首先对每一条客户机二进制指令进行检查，当监测到系统调用指令如arm的svc指令(0xef000000)时，根据当前VCPU的寄存器

参数判断系统调用类型，若属于需要监测的系统调用类型，则收集本次系统调用的参数，并触发AFL-SPY插件中编写的系统调用回调函数。

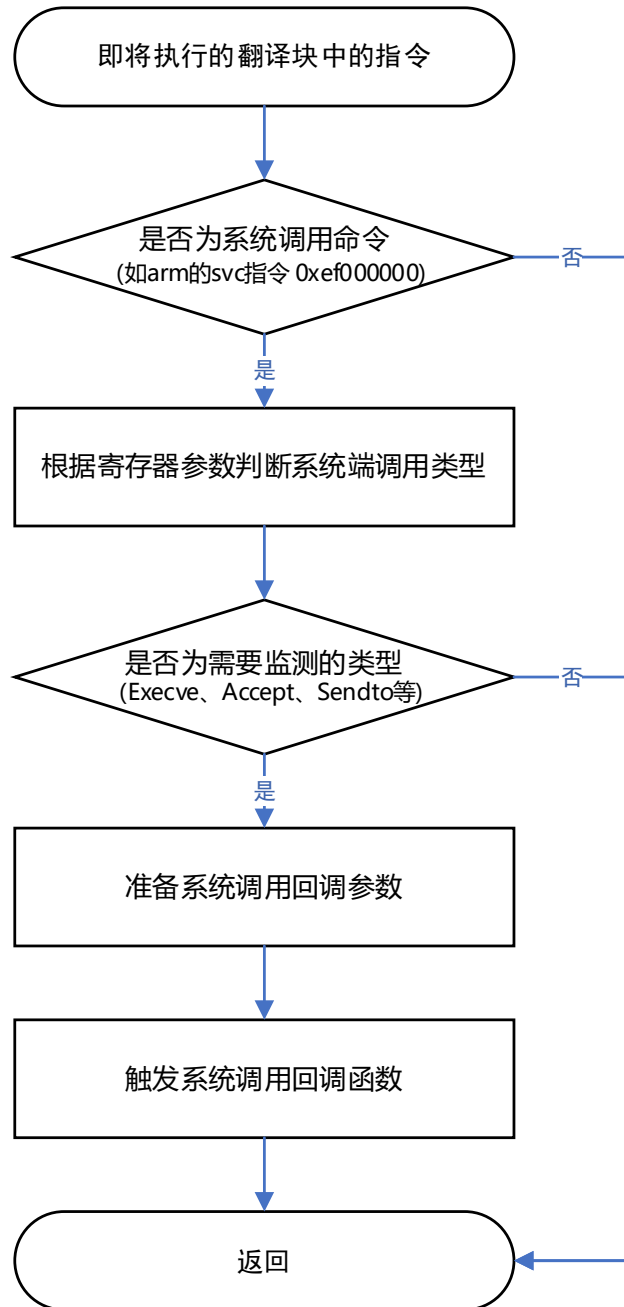


图 4-2 系统调用监测流程图

(4) 当监测到特定事件时，触发AFL-SPY中的回调函数。这一功能，由QEMU官方提供的插件机制来保障实现。本文保持该机制的框架不变，额外添加了所需的TB_EXEC_SPY和SYSCALL_SPY等事件及配套的中间函数。

(5) 翻译块执行回调：AFL-SPY插件中编写的翻译块执行回调函数(tb_exec_spy)，用于实现代码执行信息的统计。具体实现逻辑如下图所示，启动TbInfo为作为参数传给回调函数的结构体，包含当前页目录地址ctx，和当前翻译块中对应的客户机二进制代码块中第一条指令的地址pc。首先判断trace_enabled是否处于激活状态，若激活且当前进程是目标进程，则进一步检查pc的值是否处于监测范围，若处于则进行记录，更新trace_bits。

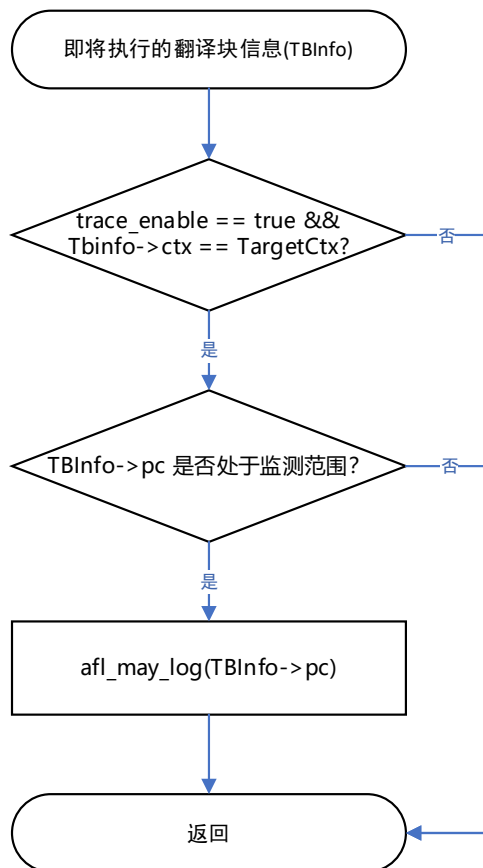


图 4-4 “翻译块执行事件”回调函数流程图

(6) 系统调用回调：AFL-SPY插件中编写的翻译块执行回调函数(syscall_spy)，用于调控模糊测试流程并和afl-fuzz进程保持同步。系统调用会调用函数的逻辑较为复杂，在下一节中进行详细讲解。

4.3 系统调用回调

AFL-SPY插件中实现的系统调用回调函数部分是调控模糊测试流程，并和afl-fuzz进程保持同步的关键逻辑所在。本文实现的AFLNetSpy原型系统需要监测三类系统调用：①用于创建进程的Execve等系统调用；②用于接收网络请求的Accept系统调用；③用于发送网络请求的Send/Sendto/Sendmsg等系统调用。下面分别进行介绍：

(1) 系统调用回调函数对Execve系统调用的处理逻辑如下图所示，当监测到Execve系统调用时，首先会提取需要创建的进程的进程名，如果进程名是用户指定的能够标识系统启动完成的进程名，则将SystemStarted设为True，表示系统启动完成，同时认为目标进程和SpyAgent进程也已经启动完成。

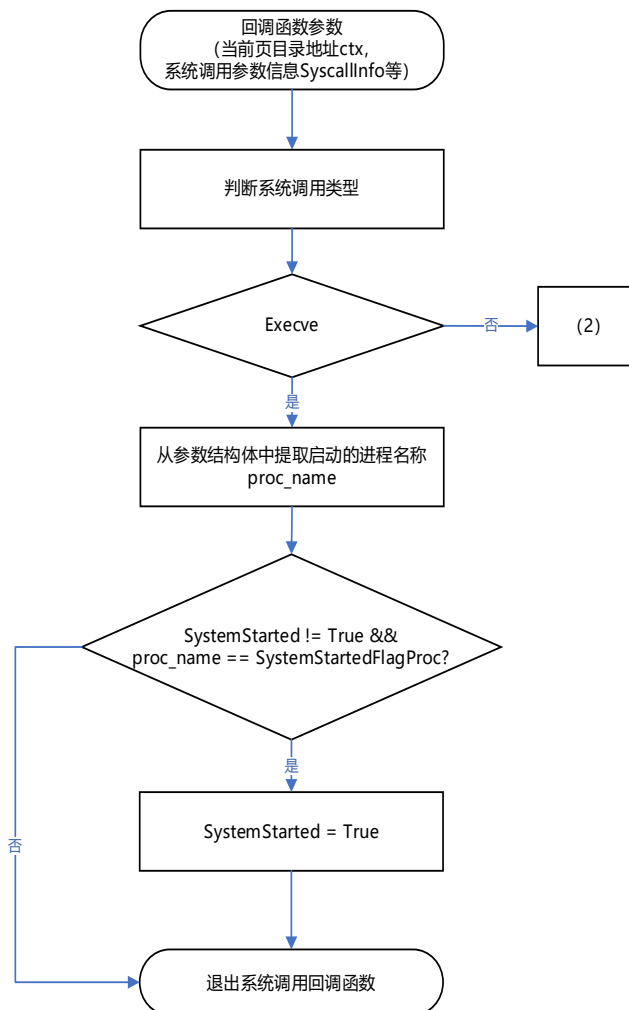


图 4-5 “Execve 系统调用”回调函数流程图

(2) 系统调用回调函数对Accept系统调用的处理逻辑如下图所示，当监测到Accept系统调用时，根据当前状态有四条执行路径，其中后三条执行路径以系统启动完成为前提：①系统启动完成前，忽略Accept调用，不进行任何操作；②设置AgentCtx：如果AgentCtx仍为初始值0，表明本次处理的请求是TEST_AGENT请求，因此把AgentCtx的值设为当前执行进程及SpyAgent进程的页目录地址ctx，并通过管道将该值返回给afl-fuzz进程实现同步；③设置TargetCtx：如果AgentCtx的值非0，而TargetCtx的值为0，表明本次处理的请求是TEST_ALIVE请求，因此把TargetCtx的值设为当前执行进程及目标测试进程的页目录地址ctx，并通过管道将该值返回给afl-fuzz进程实现同步；④重置TargetCtx：如果AgentCtx和TargetCtx的值都非0，表明本次处理的请求是RESTART_TARGET请求，因此把TargetCtx的值重置为0，表示需要重新通过TEST_ALIVE请求确定TargetCtx。

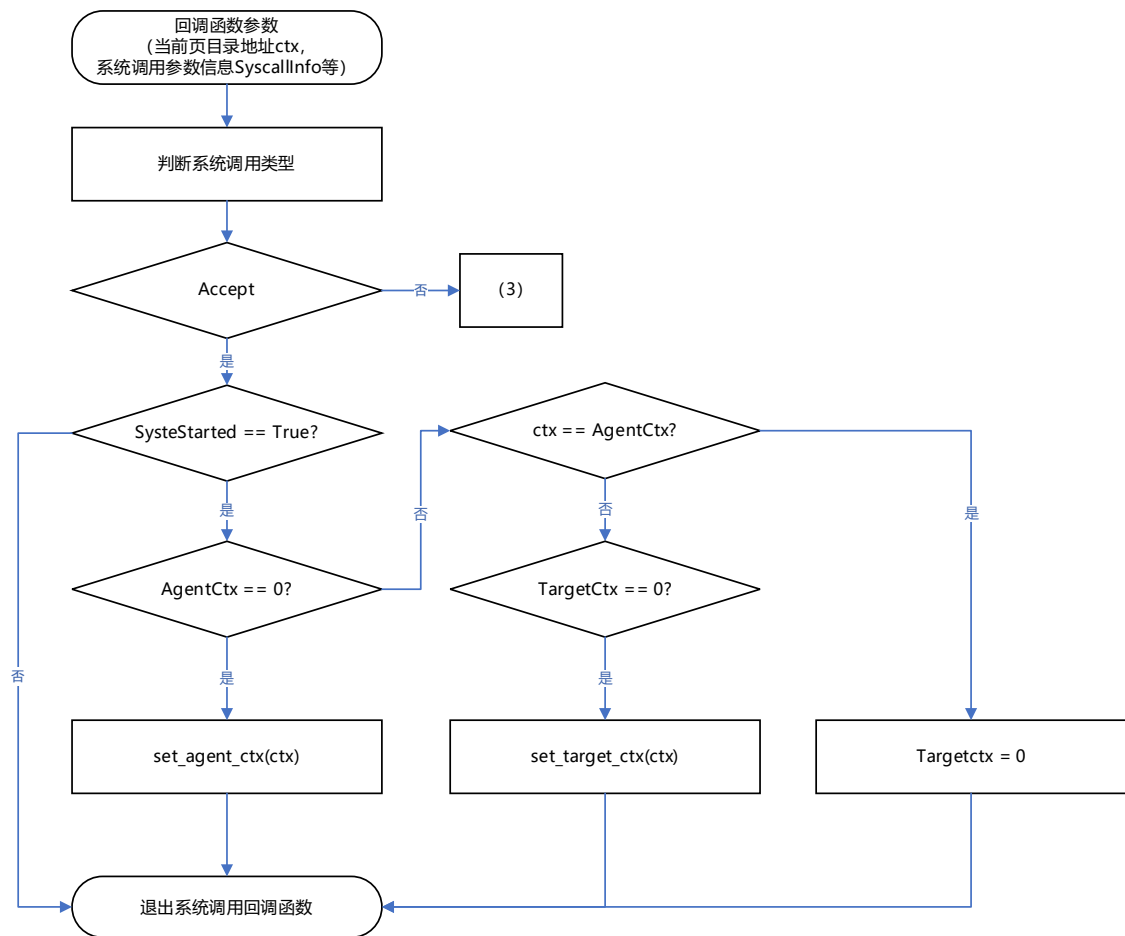


图 4-6 “Accept 系统调用”回调函数流程图

(3) 系统调用回调函数对Send/SendTo/Sendmsg等测试进程用来返回请求响应的系统调用的处理逻辑如下图所示，当监测到Send/SendTo/Sendmsg等系统调用时，首先会检查系统同是已经启动，若系统已经启动，且当前执行进程是目标进程，则进一步检查trace_enabled和next_step两个变量的值，若trace_enabled的值为1且next_step的值为0，则说明当前目标进程恰好处理完测试请求的时间点，于是把next_step的值设为0，以通知afl-fuzz进程可以开始准备下一次测试了。

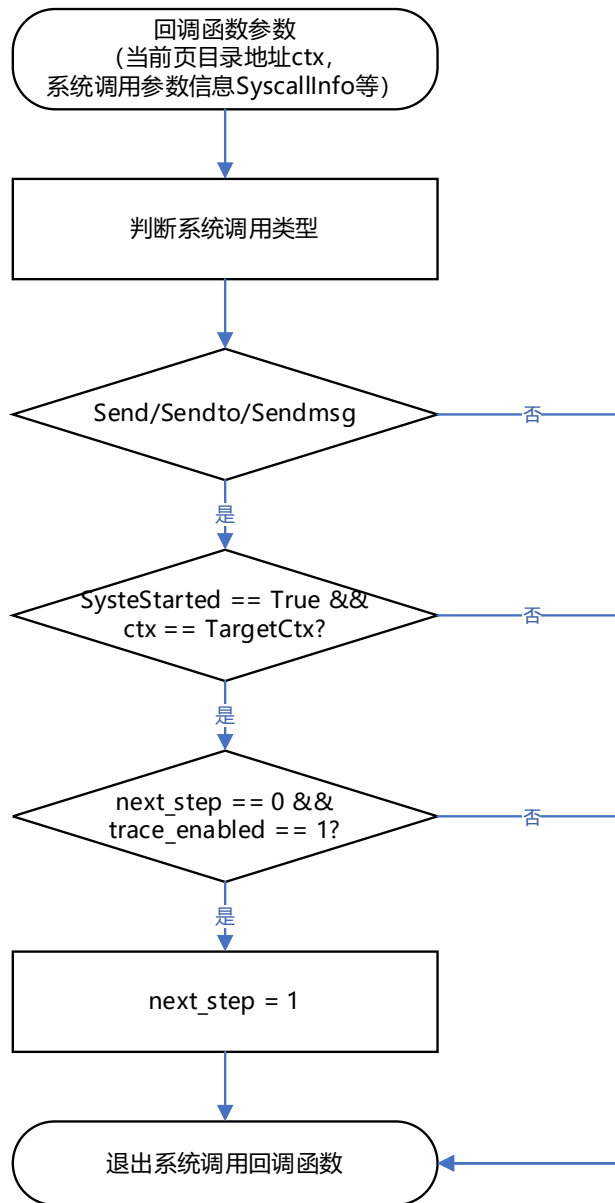


图 4-7 “Send/Sendto/Sendmsg”等系统调用”回调函数流程图

第5章 实验分析

本章首先介绍实验前的环境准备工作，然后从稳定性和性能两个方面对AFLNetSpy原型系统进行实验分析。

5.1 环境准备

在开始实验前，需要进行一些准备工作：

- (1) 准备AFLNet：由于AFLNet继承自AFL，其使用的QEMU版本为较老的2.10.0。经测试该版本的QEMU及AFL提供的修改，无法在Ubuntu22.04系统上成功进行编译。因此要想使用AFLNet的动态插桩模式，首先需要对AFLNet使用的QEMU进行修改。本文成功将AFL为QEMU-2.10.0版本提供的修改移植到了较新的8.2.0版本，并通过调试验证其运行的正确性。
- (2) 准备AFLNetSpy：根据待测固件的架构，设置QEMU-SPY系统的编译参数，并验证AFL-SPY插件libaflspy.so的可用性和正确性。
- (3) 准备固件镜像：选择的固件镜像需要保证能够使用QEMU-SPY系统成功模拟启动。本文选择使用开源固件系统OpenBmc的比较成熟的romulus版本进行后续实验，由于OpenBmc采用yocto构建系统，因此对磁盘空间较高要求且编译耗时较长。
- (4) 准备测试程序：本文使用跨平台网络框架CROW编写了一个简单二进制网络程序hello-crow.cpp。共需编译三个版本：①使用afl-g++编译测试程序，供AFLNet的静态模式使用；②使用arm-linux-gnueabi-g++编译测试程序，供AFLNet的动态插桩模式即QEMU-USER模式使用；③向OpenBmc项目中为测试程序添加一个recipe，供AFLNetSpy测试使用。
- (5) 编写请求脚本：AFLNetSpy系统要求用户根据测试目标的具体情况，编写三个请求脚本：①负责执行TEST_ALIVE请求的test_alive.sh；②负责执行TEST_AGENT请求的test_agent.sh；③负责执行RESTART_TARGET请求的restart_target.sh。三个请求脚本的具体内容见代码仓库。

5.2 稳定性分析

AFLNetSpy系统的稳定性，取决于QEMU-SPY系统得到的目标进程的代码执行信息的稳定性。本节通过比较和分析AFLNet的静态插桩模式、AFLNet的动态插桩即

QEMU-USER模式和AFLNetSpy系统模式三种情况下，针对测试程序发送相同测试请求得到的trace_bits数据的波动性，来说明AFLNetSpy系统的稳定性。

比较两个trace_bits数组相似度的算法如下：

```
输入： trace_bits_1 和 trace_bits_2
1 获取 trace_bits_1 的所有非 0 值的下标构成集合 keys1
2 获取 trace_bits_2 的所有非 0 值的下标构成集合 keys2
3 取 keys1 和 keys2 的并集,得集合 keys
4 count = 0
5 for key in keys:
6     if (key in keys1) and (key in keys2) and (trace_bits_1[key] == trace_bits_2[key]):
7         count += 1
8 return count / len(keys)
```

当两个trace_bits数组的相似度超过指定的相似度阈值时，认为二者互为对方的“家人(family)”，同时每个trace_bits也是其自身的“家人”。后续实验将通过设定不同的相似度阈值，以“家人数目(family_count)”为指标，分析三种情况下，相同测试请求对应的多个trace_bits的波动性和稳定性。

(1) AFLNet的静态插桩模式下，100次请求生成的100个trace_bits(实际为110次请求筛去前10项不稳定数据)的“家人数目(Family_Count)”的统计结果如下(三次实验取均值)：

表 5-1 AFLNet 静态插桩模式下的 Family_Count 统计结果

相似度阈值	AFLNet 静态插桩模式: Family_Count				
	最大值	最小值	平均值	中位数	众数
0.9	100.00	100.00	100.00	100.00	100.00
0.95	100.00	100.00	100.00	100.00	100.00
0.99	100.00	100.00	100.00	100.00	100.00
0.999	99.00	34.00	98.03	99.00	99.00
0.9999	3.00	1.00	1.38	1.00	1.00

(2) AFLNet的动态插桩模式即QEMU-USER模式下，100次请求生成的100个trace_bits(实际为110次请求筛去前10项不稳定数据)的“家人数目(Family_Count)”的统计结果如下(三次实验取均值)：

表 5-2 AFLNet 动态插桩模式下的 Family_Count 统计结果

相似度阈值	AFLNet 动态插桩模式: Family_Count				
	最大值	最小值	平均值	中位数	众数
0.9	100.00	100.00	100.00	100.00	100.00
0.95	100.00	100.00	100.00	100.00	100.00
0.99	96.00	3.33	93.53	96.67	96.67
0.999	89.67	1.33	80.08	89.33	89.33
0.9999	3.00	1.00	1.26	1.00	1.00

(3) AFLNetSpy系统模式下，100次请求生成的100个trace_bits(实际为110次请求筛去前10项不稳定数据),的“家人数目(Family_Count)”的统计结果如下(三次实验取均值):

图 5-3 AFLNetSpy 系统模式下的 Family_Count 统计结果

相似度阈值	AFLNetSpy: Family_Count				
	最大值	最小值	平均值	中位数	众数
0.5	100.00	100.00	100.00	100.00	100.00
0.6	99.00	1.00	98.02	99.00	99.00
0.7	99.00	1.00	97.48	99.00	99.00
0.8	97.33	1.00	81.90	89.33	89.33
0.9	81.33	1.00	65.59	80.67	80.67
0.95	71.67	1.00	50.61	70.00	70.33
0.99	61.33	1.00	38.32	59.83	61.00
0.999	23.33	1.00	10.61	10.00	1.00
0.9999	23.33	1.00	10.60	10.00	1.00

根据以上实验结果，可以看到三种情况下的实验都表明，对于网络应用来说，相同的测试用例，得到的代码执行信息并不一定完全相同，可能存在波动情况。根据AFL白皮书的解释，这可能是由于网络应用的多线程或轮询机制导致的，属于正常现象。

同时，还可以发现AFLNetSpy系统收集到的trace_bits数据的稳定性要比AFLNet略差一些，原因主要有两点：①AFLNet每次测试会重启测试进程，而AFLNet会持续使用同一个测试进程直到该进程崩溃；②AFLNetSpy会记录测试进程在存活期间执行的所有指令信息，而AFLNetSpy只会记录测试进程接收请求(Accept)到返回响应(Send/Sendto/Sendmsg)这一过程中执行的指令信息。

5.3 性能分析

本节通过设计实验，分别在AFLNet的静态插桩模式、AFLNet的动态插桩即QEMU-USER模式和AFLNetSpy系统模式三种情况下运行300次测试(实际运行310次筛去前10项不稳定数据)，收集得到300项速率数据，进而比较三种情况的执行效率。实验结果如下(三次实验取均值)：

图 5-4 三种模式执行速率的对比数据

模式	每秒执行次数			
	最大值	最小值	平均值	中位数
AFLNet 静态插桩模式	33.30	23.70	31.76	32.14
AFLNet 动态插桩模式	7.29	6.40	6.94	6.95
AFLNetSpy 系统模式	19.57	14.50	17.05	17.01

根据以上实验结果，可以看出：①AFLNet静态插桩模式的效率要远高于动态插桩模式。略高于AFLNetSpy，这主要是因为QEMU的二进制翻译过程带来了较大的时间消耗；②AFLNetSpy的执行速率约为AFLNet动态插桩模式的2.5倍，主要原因为AFLNet每次测试前都会重启一个新的测试进程，而AFLNetSpy会持续利用同一个测试进程直到该进程崩溃，从而节省时间提高执行效率。

结 论

本文提出了一个新颖的模糊测试框架AFLNetSpy，它基于AFLNet，通过结合QEMU的AFL-SPY插件，能够实现对固件中网络应用的灰盒模糊测试。AFLNetSpy的主要创新点包括：①通过QEMU-SPY监测客户机系统中目标进程和网络相关的系统调用，精准识别目标进程处理测试请求的时间区间；②结合QEMU的HELPER机制和插件机制实现客户机系统中目标进程的标识，并进一步实现目标进程代码执行信息的收集。另外，QEMU-SPY的关键功能以QEMU插件的形式实现，即AFL-SPY插件(libaflspy.so)，使得系统具有较高的可扩展性。

本文首先介绍了模糊测试领域的研究概况，并对和本文工作相关的几项工作的特点和不足进行了较为详细的说明。然后在第三章阐述了针对固件内网络应用进行灰盒模糊测试的关键问题和解决思路，也就是AFLNetSpy系统的设计要点。随后，在第四章对AFLNetSpy系统的整体架构和实现细节进行了详细的讲解。最后，在第五章从稳定性和性能两个方面出发进行实验设计，将AFLNetSpy系统和AFLNet的两种模式进行比较分析，从而成功验证AFLNetSpy系统的可用性。

综上，本文提出的针对固件内网络应用进行灰盒模糊测试的方法，以及开发出的AFLNetSpy原型系统，将灰盒模糊测试方法成功拓展到了固件网络应用领域，对于促进相关领域的研究具有重要意义。同时，本文的工作还存在一定不足，如局限于固件内的网络应用而难以泛化到其他类型的二进制应用、效率还有提升空间等，有待后续进一步研究。本文开发的AFLNetSpy系统及相关代码，可访问<https://github.com/czxvan/AFLNetSpy> 代码仓库进行查看。

参考文献

- [1] Miller B P, Fredriksen L, So B. An empirical study of the reliability of UNIX utilities[J]. Communications of the ACM, 1990, 33(12): 32–44.
- [2] Zalewski M. American Fuzzy Lop[J]. 2016. <https://lcamtuf.coredump.cx/afl/>.
- [3] Böhme M, Pham V-T, Roychoudhury A. Coverage-based Greybox Fuzzing as Markov Chain[A]. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security[C]. Vienna Austria: ACM, 2016: 1032–1043.
- [4] Lyu C, Ji S, Zhang C, et al. MOPT: Optimized Mutation Scheduling for Fuzzers[J].
- [5] Circumventing Fuzzing Roadblocks with Compiler Transformations[J]. 2016. <https://lafintel.wordpress.com/2016/08/15/circumventing-fuzzing-roadblocks-with-compiler-transformations/>.
- [6] Aschermann C, Schumilo S, Blazytko T, et al. REDQUEEN: Fuzzing with Input-to-State Correspondence[A]. Proceedings 2019 Network and Distributed System Security Symposium[C]. San Diego, CA: Internet Society, 2019.
- [7] Schumilo S, Aschermann C, Gawlik R, et al. kAFL: Hardware-Assisted Feedback Fuzzing for OS Kernels[J].
- [8] Michael Eddington. Peach fuzzing platform[J]. <https://peachtech.gitlab.io/peach-fuzzer-community/>.
- [9] Pham V-T, Boehme M, Santosa A E, et al. Smart Greybox Fuzzing[J]. IEEE Transactions on Software Engineering, 2020: 1–1.
- [10] Fioraldi A, Maier D, Eißfeldt H, et al. AFL++: Combining Incremental Steps of Fuzzing Research[J].
- [11] Fioraldi A, Maier D C, Zhang D, et al. LibAFL: A Framework to Build Modular and Reusable Fuzzers[A]. Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security[C]. New York, NY, USA: Association for Computing Machinery, 2022: 1051–1065.
- [12] Pham V-T, Bohme M, Roychoudhury A. AFLNET: A Greybox Fuzzer for Network Protocols[A]. 2020 IEEE 13th International Conference on Software Testing, Validation and Verification (ICST)[C]. Porto, Portugal: IEEE, 2020: 460–465.
- [13] Andronidis A, Cadar C. SnapFuzz: high-throughput fuzzing of network applications[A]. Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis[C]. Virtual South Korea: ACM, 2022: 340–351.
- [14] Hertz J, Newsham T. Project Triforce: Run AFL On Everything[J].
- [15] Bellard F. QEMU, a Fast and Portable Dynamic Translator[J]. 2005.
- [16] Zheng Y, Davanian A, Yin H, et al. {FIRM-AFL}: {High-Throughput} Greybox Fuzzing of {IoT} Firmware via Augmented Process Emulation[A]. 2019: 1099–1114.
- [17] Zheng Y, Li Y, Zhang C, et al. Efficient greybox fuzzing of applications in Linux-based IoT devices via enhanced user-mode emulation[A]. Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis[C]. Virtual South Korea: ACM, 2022: 417–428.
- [18] Meng R, Mirchev M, Böhme M, et al. Large Language Model guided Protocol Fuzzing[A]. Proceedings 2024 Network and Distributed System Security Symposium[C]. San Diego, CA, USA: Internet Society, 2024.

- [19] Henderson A, Prakash A, Yan L K, et al. Make it work, make it right, make it fast: building a platform-neutral whole-system dynamic binary analysis platform[A]. Proceedings of the 2014 International Symposium on Software Testing and Analysis[C]. San Jose CA USA: ACM, 2014: 248–258.
- [20] Dovgalyuk P, Fursova N, Vasiliev I, et al. QEMU-based framework for non-intrusive virtual machine instrumentation and introspection[A]. Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering[C]. Paderborn Germany: ACM, 2017: 944–948.

致 谢

本科生涯即将结束，回顾过去四年，可以看到自己在能力上的提升和性格上的改变，这是四年北理时光为我留下的印记。

首先，感谢谭毓安老师对本次毕业设计中几个关键点的建议和指导，感谢实验室的师兄和同学提供的帮助，感谢我的家人们对我的鼓励和支持。

然后，感谢一起参与各项竞赛的队友以及网安俱乐部的朋友们，和你们一起成长进步讨论交流，让我受益匪浅。感谢我的舍友们，四年里尤其是最后一年我们一起度过了大部分的时光，对此我感到十分幸运和愉快。愿友谊长青，时常联系。

最后，祝大学四年里遇到的老师们身体健康，工作顺利。祝我的同学朋友们前程似锦，后会有期。