一种面向异构固件的高效靶向分析技术

江子锐，梁辰，张子龙，王奕森*

（信息工程大学网络空间安全学院，河南 郑州 450001）

摘　要：嵌入式设备的快速增长和广泛应用，带来便利的同时也引入了巨大的安全风险，其中，固件安全是关键风险点之一。嵌入式设备固件数量多、格式复杂，且很多经过加密、混淆，使安全分析人员难以快速有效地解析固件并发现隐藏脆弱点。针对以上问题，提出一种面向异构固件的高效靶向分析技术。首先，研究多粒度分析方法、文件自动分类、关键信息输出和靶向提取等技术，实现深度可控的固件靶向解析；然后，建立文件系统特征库，研究基于特征值匹配的靶向识别技术，增强混淆固件的识别能力，扩展文件系统识别范围；最后，设计爬虫提取不同厂商固件并构建万级固件库作为基础支撑，实现基于邻近版本的固件靶向解密。设计实现了 FTA 自动化固件解析系统并进行测试，实验结果表明，与主流固件分析工具 Binwalk 相比，FTA 实现的多粒度分析方法对固件的解析速度平均提升 42.59%，优化输出模式实现固件中目标文件的靶向提取，扩展了对多种文件系统特征值的识别能力，为嵌入式系统安全领域中的固件解析工作提供了有力支持。

关键词：嵌入式设备安全；固件自动解析；固件加密解密；固件靶向分析；大规模固件分析

中图分类号：TP311.5　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0069914

An Efficient Targeting Analysis Technique for Heterogeneous Firmware

JIANG Zirui, LIANG Chen, ZHANG Zilong, WANG Yisen*

(School of Cybersecurity, Information Engineering University, Zhengzhou 450001, Henan, China)

【Abstract】The widespread use and diversification of embedded devices have introduced unparalleled convenience and formidable security vulnerabilities, particularly in firmware security. The intricate nature of embedded device firmware, coupled with its sheer volume and the adoption of encryption and obfuscation techniques, presents a formidable challenge for security analysts seeking to uncover hidden vulnerabilities efficiently. In response to this challenge, this study proposes an innovative targeting analysis technique customized to heterogeneous firmware. First, the study explores multi-granularity analysis methods, automatic document categorization, key information extraction, and target delineation techniques to enable nuanced and depth-controllable firmware analysis. Next, it establishes a comprehensive file system feature library and introduces a novel target recognition approach based on eigenvalue matching, enhancing the discernment capabilities for obscure firmware and expanding the breadth of file system identification. Furthermore, the study develops a specialized crawler to procure firmware from diverse vendors, leading to the construction of a 10 000-level firmware library that is crucial for targeted decryption based on neighboring versions. An FTA automated firmware parsing system is conceptualized and empirically validated, showing significant enhancements over mainstream firmware analysis tools such as Binwalk. Specifically, FTA′s multi-granular analysis method elevates the firmware parsing speed by an average of 42.59%, whereas the optimized output mode facilitates targeted file extraction and extends recognition capabilities across multiple file system feature values. FTA provides robust support for firmware parsing within the domain of embedded system security.

【Key words】embedded device security; firmware automation parsing; firmware encryption and decryption; firmware targeting analysis; large-scale firmware analysis

0　引言

近年来，物联网技术正在加速向各行各业渗透。根据《物联网操作系统安全白皮书》[1]显示：全球物联网连接数保持高速增长，2020 年全球物联网总连接数达到 131 亿，预计到 2025 年，连接规模将达到 246 亿。随着物联网技术的普及和应用，越来越多的物联网智能设备被部署运用于日常生活和关键基础设施之中。然而物联网这项新兴技术在改善生活、提供便利的同时，也带来了诸多新的安全问

收稿日期：2024-05-24　　**修回日期：**2024-06-29

基金项目：河南省重点研发专项(221111210300)。

通信作者 E-mail：* 851067568@qq.com

题[2]，例如：2023 年 5 月 19 日，Wemo Mini 智能插头被曝出存在 CVE-2023-27217 安全漏洞，设备一旦入网，黑客便可以控制插入该设备的任何电子设备来远程操作；SonicWall Capture 实验室的网络威胁研究人员于 2022 年记录了 1.123 亿个物联网恶意软件攻击实例，与 2021 年相比增加了 87%。设备的安全问题主要体现在固件中，固件解析是固件漏洞挖掘的前序阶段[3]，为固件的安全性分析提供基础支持，因此，设备固件分析十分重要，尤其是大规模的固件分析。2014 年，ANDREI 等[4]提出了一个大规模的固件分析工具，在没有进行复杂静态分析的情况下就在超过 693 个固件图像中发现了 38 个未知的安全漏洞，此外，通过关联不同的固件中的相似文件，将其中一些漏洞扩展到超过 12 个不同的设备，其中一些漏洞影响了至少 14 万台联网设备。

以上现状说明大规模固件分析能获得较好的效果，但同时也引出了一个新的问题：开展大规模自动化固件分析时，目前常用解包的 Binwalk[5]、BAP[6]和 BARF[7]等工具存在准确率低、解密固件难和对批量固件分析时功能较为单一的问题。为加强对嵌入式固件的解包能力，HAIDER 等[8]提出了一种对嵌入式固件进行解包[9]和解密[10]的分析框架——DUDE，该框架对特定厂商的固件文件系统[11]解包过程进行了优化，但仍存在解包不全、加密固件种类支持少等不足。针对当前研究存在的问题，研究多种嵌入式固件提取和解包技术[12]，提出一种面向异构固件的高效靶向分析技术，设计 FTA 固件靶向解析系统，通过固件靶向解析技术提升固件安全分析的效率。本文的主要工作如下：

1）针对当前固件解析工具定位和提取固件目标文件（如网络服务文件、配置文件等）准确率低、速度慢等问题，提出固件多粒度分析方法，通过浅度匹配解析和递归深度解析实现批量固件快速解析，提取操作系统内核以及文件系统，对目标文件安全扫描，发现弱口令、可疑后门[13]（如授权密钥文件、硬编码的证书以及 Web 服务器配置文件）等，通过以上信息实现文件自动分类和关键信息输出，最后完成目标文件靶向提取，为后续判断固件中是否存在配置风险[14]、漏洞隐患[15]、密钥安全、隐私信息泄露[16]、代码安全等提供支撑。

2）针对当前固件解析工具对文件系统支持种类少、解析效果不理想等问题，提出固件靶向分析技术，扩展对文件系统格式[17]的支持并实现自动化靶向固件解析。分析 JFFS2（Journalling Flash FileSystem v2）、YAFFS（Yet Another Flash File System）、Cramfs（Compressed ROM file system）、Romfs（ROM file system）、RamDisk、Ramfs（RAM file system）等文件系统格式，扩充对不同文件系统的支持。

3）针对当前研究缺少工具实现不同类型和版本加密固件分析和解密的问题，研究固件安全加密技术[18]，提出基于邻近版本的固件靶向解密方法，通过爬虫、抓包和硬件接口[19]提取等方法获取分析固件[20]，并构建固件库作为基础，实现基于邻近版本的固件靶向解密。

4）为验证所提技术的先进性及有效性，与经典固件分析工具 Binwalk 在功能和解析速度上进行对比实验。在批量固件解析方面，所提方法较 Binwalk 速度平均提升 42.59%，能够靶向识别 2 类文件系统的 9 个不同特征值，并且新增靶向提取、关键信息输出和自动分类等功能，显著提升了大规模固件安全分析的效率。

1 相关技术

固件分析是嵌入式系统安全分析[21]的关键，可以分为动态分析（如 Rehosting 托管技术和 Fuzzing 模糊测试等）和静态分析。

动态分析主要利用经典工具 Firmadyne[22]来提供模拟真实硬件/外设的仿真环境进行固件仿真，KIM 等[23]进一步研究了大规模动态分析实现的基于启发式方法的仿真工具 FirmAE。但由于设备外设多导致固件仿真成功率低，严重限制了固件动态分析技术的发展。

静态分析通过分析固件中的 ASCII 字符串或者二进制数据，使用 IDA（Interactive Disassembler Professional）等反汇编工具来获取固件函数等有效信息进行相似性分析[24]并发掘漏洞[25]，如 PEWNY 等[26]提出的二进制级推导已知漏洞的签名技术。设备厂商在开发设备固件时将固件存储在 EEPROM 芯片[27]中，并采用不同的压缩方式和文件系统，如 rar、tar、bin 等压缩方式和 Squashfs、YAFFS、JFFS2 等文件系统，主流的格式可以被当前固件分析工具识别和分析。

随着厂商对安全性的重视程度的提高，会使用定制化的文件系统，如 UBI、coreboot 等，并且为了防止固件被黑客恶意攻击，也会加入防分析措施，主流的防分析方法有 2 种：第 1 种方法是对老版本固件升级时引入加解密算法的程序，对固件进行升级加密保护；第 2 种方法是在固件的文件系统头部特

征值[28]处进行混淆[29]，导致分析工具无法识别文件系统的特征值而分析失败。

分析人员可通过固件分析工具[30]手动分析，然而加密固件和混淆特征值等防分析方法导致手动分析效率低，难以适应大规模固件分析的需要。

本文提出一种固件靶向分析技术，研究固件多粒度分析、关键目标靶向提取和可扩展文件系统提取等关键技术，实现对固件格式的准确解析，并构建固件库，实现基于邻近版本的固件靶向解密。

2 固件多粒度分析及靶向提取技术

2.1 FTA 自动化固件解析工具系统架构

针对当前固件解析工具在固件解析时存在功能较单一、解析信息输出方式少、解析深度不可控等问题，提出一种固件靶向分析技术，设计实现了 FTA 自动化固件解析系统，系统框架如图 1 所示。

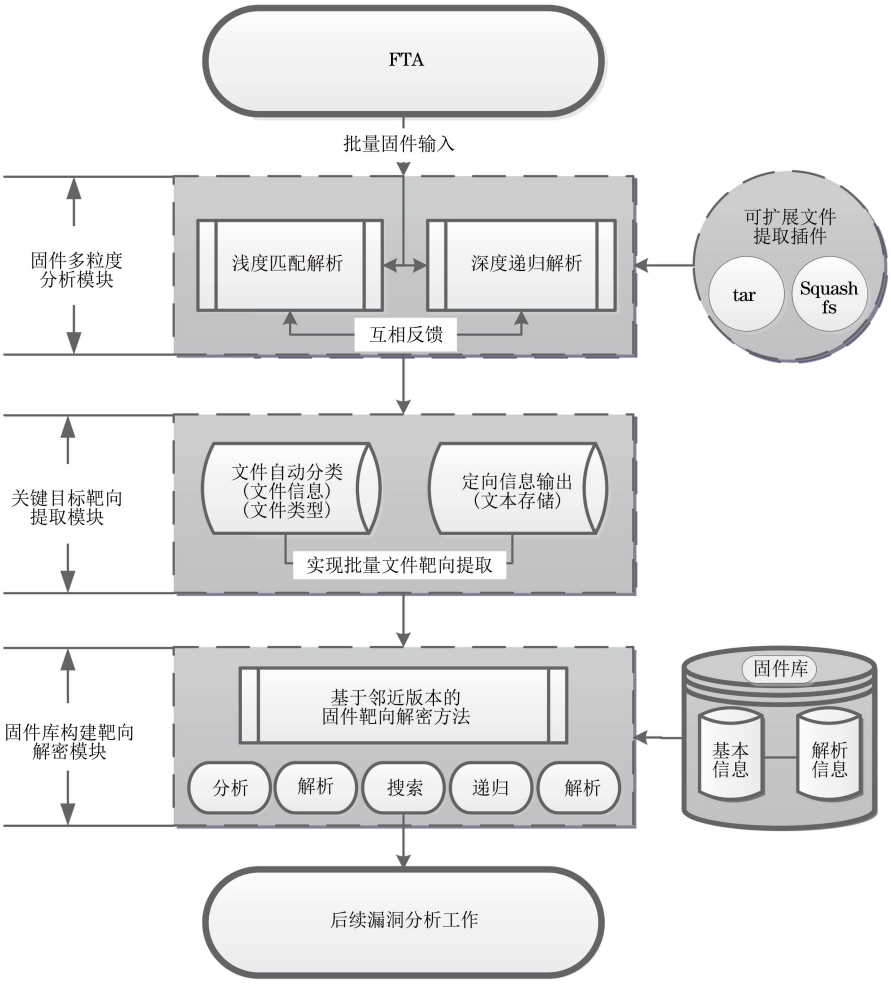

图 1 FTA 自动化固件解析系统框架

Fig. 1 FTA automated firmware parsing system framework

FTA 自动化固件解析工具的系统架构主要分为 3 个主要功能模块：

1)固件多粒度分析模块。在批量固件输入后，通过浅度匹配解析和深度递归解析对不同压缩格式的文件进行解包，其中深度递归解析模块通过读取并匹配固件的特征值，以识别目标设备固件的压缩方式。确定压缩格式后进行靶向解压，扫描解压后的所有文件类型，对仍然存在的压缩文件递归循环解压，直至不能继续解压。浅度匹配解析对文件类型预设黑/白名单，对白名单中的文件进行基础解析，获取固件基本信息，同时清除黑名单中的冗余文件，把剩余压缩文件转入深度递归解析。在多层解析中对预设黑/白名单强化反馈，优化解析效果。

2)固件靶向提取模块。为增强 FTA 对不同文件的识别能力，设计靶向提取插件，对混淆特征值的 tar 文件和 Squashfs 文件系统实现靶向识别。通过分析模块靶向解包后，获取固件字节序、Magic 值、CRC 校验码等关键信息。靶向提取模块包含文件自动分类和定向信息输出功能。通过解析后文件的后缀类型和包含的敏感信息两方面分类，将分析后固件的 MD5 码、特征值等信息以 json 文件输出。最后根据分类信息检索关键词实现靶向提取，以便快速

获得目标文件用于后续的函数分析和漏洞挖掘。

3)固件靶向解密模块。针对加密固件,设计基于邻近版本的靶向解密方法。由靶向提取获取的信息构建数据库,其中包含对固件名切分截取的固件基本信息,以及在自动分类和信息输出中提取的解析信息。在预设固件库的基础上,通过邻近版本固件靶向解密方法实现对固件的解密,进一步提高自动解析工具的分析性能。

2.2 固件多粒度分析方法

2.2.1 固件结构分析和压缩格式识别方法

大规模固件解析中存在许多不同种类、大小的固件,如果只对固件简单解析会导致提取信息不全面;如果对所有固件全部解压,则会解包大量冗余文件,造成磁盘空间浪费。为提升大规模固件分析的效率和准确率,实现解析深度可控的自动化批量固件解析,提出固件多粒度分析方法。

针对固件结构识别困难导致无法准确解析内部文件的问题,开展固件结构分析方法研究,实现针对固件头(Firmware Header)、引导加载程序(Boot Loader)、操作系统内核(Kernel)、根文件系统(RootFs)以及其他应用配置文件的准确解析,固件结构分析如图 2 所示。

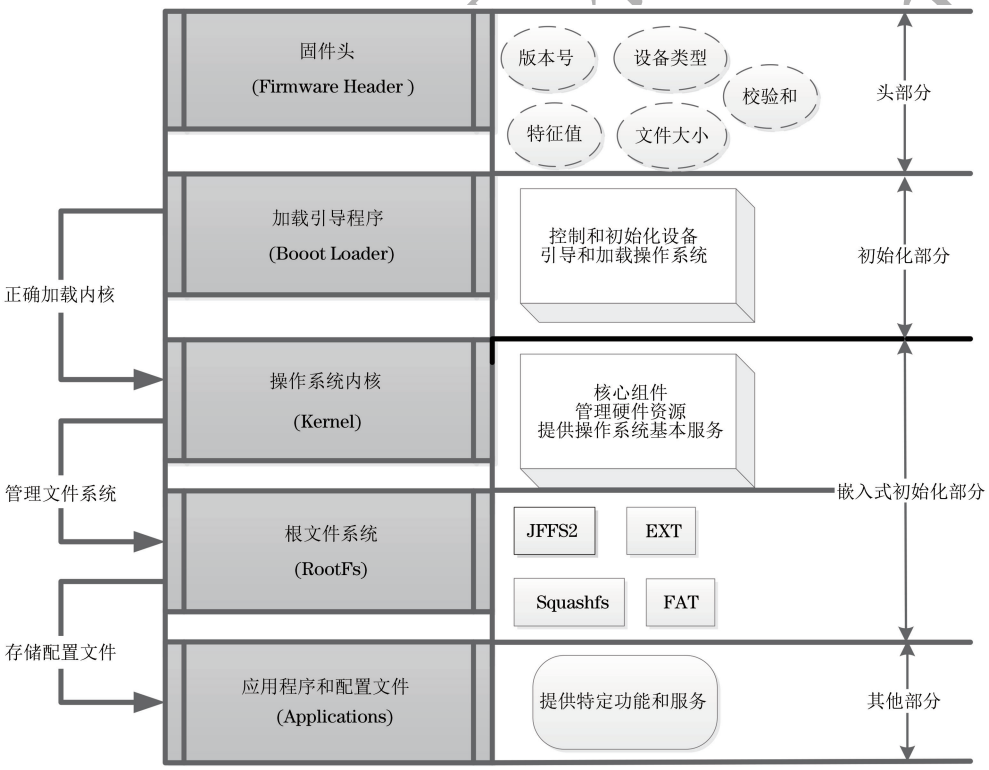

图 2 固件组成结构

Fig. 2 Firmware composition structure

嵌入式设备的处理器架构、操作系统和文件系统具有多样性特征,例如:网络设备多为 MIPS 架构、Linux 内核和 Squashfs 文件系统;智能终端多为 RISC-V 架构、FreeRTOS 内核和 YAFFS 文件系统;高性能通信设备多为 ARM 架构、VxWorks 内核。为了实现异构固件的自动化分析,对嵌入式系统启动、内核以及文件系统展开研究。

首先,启动模块(加载引导程序 Boot Loader)作为固件初始化部分,是正确加载内核的基础,目前主流固件启动模块的主要信息如表 1 所示。

操作系统内核作为固件的核心组件,能有效管理复杂的系统资源。基于固件的异构特性,针对不同的功能,每种内核都有独特的优势和应用领域,主流系统内核如表 2 所示。

表 1 开放源码的 Linux 引导程序

Table 1 Open source Linux boot loader

Boot Loader	描述	X86	ARM	PowerPC
LILO	Linux 磁盘引导程序	是	否	否
GRUB	GNU 的 LILO 替代程序	是	否	否
Loadlin	从 DOS 引导 Linux	是	否	否
ROLO	从 ROM 引导 Linux 不需要 BIOS	是	否	否
Etherboot	通过以太网卡启动 Linux 系统的固件	是	否	否
LinuxBIOS	完全替代 BUIS 的 Linux 引导程序	是	否	否
BLOB	LART 等硬件平台的引导程序	否	是	否
U-boot	通用引导程序	是	是	是
RedBoot	基于 eCos 的引导程序	是	是	是

表 2 主流固件内核

Table 2 Mainstream firmware kernel

操作系统内核	描述	优点
嵌入式 Linux	以 Linux 为基础的嵌入式操作系统	内核小、效率高，内核的更新速度很快
eCOS	小型即时操作系统	灵活的电源管理功能，良好的通信能力
Palm OS	专用于掌上电脑(PDA)上的一种操作系统	软硬件紧密度高，安全性强
Windows CE	微软公司嵌入式移动计算平台的基础	图形界面出色，开发效率高
VxWorks	实时操作系统(RTOS)	可靠性强，实时性好，具有高度灵活性
FreeRTOS	轻量级实时操作系统内核	移植简单，高质量管理机制

固件根据存储设备的硬件特性、系统需求采用不同文件系统。常见的文件系统格式有 YAFFS、JFFS2、Squashfs、Cramfs、MPFS、FAT、Romfs 等。主流文件系统信息如表 3 所示。

表 3 主流文件系统信息

Table 3 Mainstream file system information

文件系统	描述	特点	Magic 值
Cramfs	只读压缩文件系统	以压缩形式存储文件数据，空间特性强	0x28cd3d45
Squashfs	基于 Linux 内核使用的压缩只读文件系统	文件压缩率高，检测并删除重复文件	sqsh 等 7 种
YAFFS	NAND 型闪存日志文件系统	提供损耗平衡和掉电保护，可以有效地避免意外掉电对文件系统一致性和完整性的影响	\x00\x00\x00\x01\xFF\xFF 等 2 种
JFFS2	NOR 型闪存日志文件系统	可读写操作，具有垃圾回收机制	0x1985
Romfs	只读文件系统	节省内存，底层结构简单	-rom1fs-\0
EXT	扩展文件系统	具有良好的兼容性	0xEF53
Wind River MemFS	内存文件系统	可靠性强，常用于 VxWorks 固件	owowowowowowowo wowowowowowowow

基于固件的异构特性，FTA 通过特征值匹配实现高效固件结构分析。首先进行固件头解析，得到版本号、特征值、入口地址等关键信息。然后搜索固件头中特定字符串进行组件版本号、文件大小等靶向识别，并通过特征值匹配解析加密标记和签名。根据解析得到的特征值，循环迭代检索固件中的 Magic Bytes，查询并提取引导程序、操作系统内核和文件系统数据。

针对单层压缩固件，FTA 通过识别压缩算法签名来匹配压缩类型，定位二进制文件中压缩文件的偏移地址，并根据识别结果完成解包，但该方法对嵌套压缩文件无法完全解压。

固件中的嵌套压缩会导致无法完成固件完全解包，初步识别只能对二进制数据流进行基本分析，不能识别出最外层文件包含的压缩文件。针对此问题，提出多粒度分析方法，对于嵌套文件和单一文件分别进行解析，实现基于浅度匹配解析的固件分析和基于深度递归解析的固件分析。

2.2.2 基于浅度匹配解析的固件分析

为实现批量未嵌套压缩固件快速解析，提出浅度匹配解析方法，分析流程如图 3 所示，FTA 通过预设白名单和黑名单来对批量输入的解析文件类型进行区分。

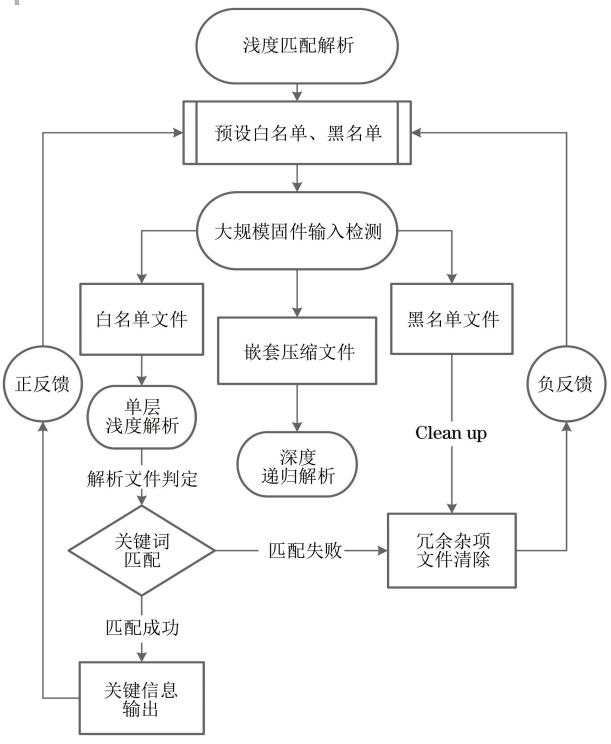

图 3 浅度匹配解析实现流程

Fig.3 Shallow matching parsing implementation flow

对可以直接解析获取相应信息的文件，如 bin、hex、elf、img、rootfs、kernel 等后缀文件加入白名单列表。

对不能解析提取到有关信息的文件，如可执行文件 x-executable、x-dosexec、x-object 以及 PDF、MS Word、Video 等与解析无关的杂项文件加入到黑名单列表。

对可能包含嵌套压缩的文件，如 zip、tar、LZMA 等无法一次提取有关信息的文件进行输出，并在后续转到深度递归解析进行提取。

批量固件分析时进入识别模块：直接对文件类型进行白、黑名单列表判断，单层浅度解析白名单内的固件文件；清除后缀在黑名单中的文件，减少占用磁盘空间；对于可能包含嵌套压缩的文件，转入深度递归解析中提取。

识别模块结束后，进入信息输出模块，对白名单文件解析后输出基础信息，将 kernel 后缀文件单独提取并保存到 Image 文件夹。然后对解析文件进行判定，通过关键词匹配完成靶向输出。匹配失败的文件列入黑名单并清除。

通过关键词匹配和输入检测结果对白名单和黑名单进行正、负反馈，强化匹配能力，逐步完善对批量固件的识别精准度，并减少冗余文件对磁盘空间的占用，实现浅度匹配解析功能。

2.2.3 基于深度递归解析的固件分析

为解决固件解包不彻底的问题，设计基于深度递归解析的固件分析方法。FTA 执行深度递归解析方法流程如图 4 所示，可以分为 3 个模块，具体如下：

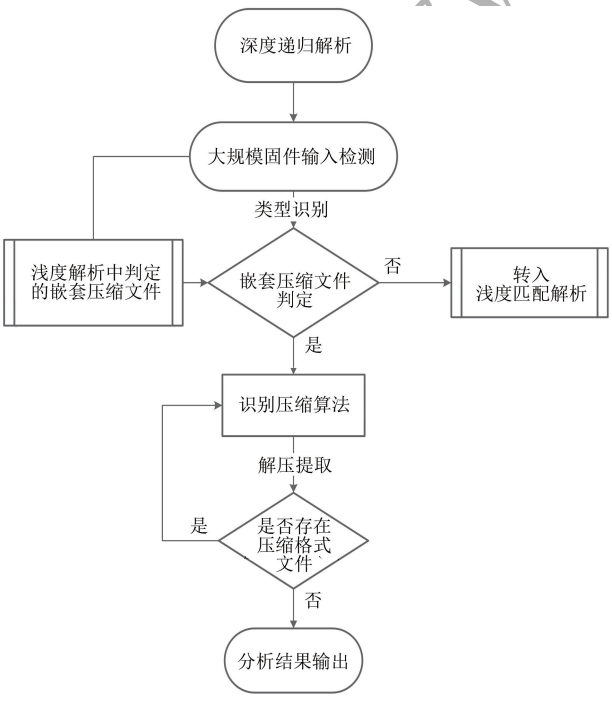

图 4 深度递归解析实现流程

Fig. 4 Deep recursive parsing implementation flow

1）识别模块。对批量固件输入进行类型识别检测，判定是否为嵌套压缩文件，如果是，则进入压缩算法识别模块；否则将此类文件转入浅度匹配解析模块。

2）压缩提取模块。对判定为嵌套压缩类型的文件和从浅度解析中转入的压缩文件进行压缩算法识别，批量扫描所有文件的压缩算法并提取。

3）迭代递归模块。初次识别提取后，对解压后的文件再次扫描，将仍存在的可压缩文件作为输入递归提取，直至只存在无法识别的部分，则停止解压，并将所有解析提取获取的信息输出，形成迭代递归的扫描和解压流程。

FTA 通过以上流程完成批量固件的深度递归解析，并输出详细解析信息以供后续逆向分析和漏洞挖掘使用。

2.3 自动分类和靶向提取方法

FTA 浅度和深度解析模块对批量固件解析提取后，得到大量目标文件，如 Config 配置文件、通用网关接口 cgi 文件、img、txt 格式文件等。针对目标文件搜索时间长、关键信息提取复杂的问题，提出关键目标靶向提取方法。FTA 的靶向提取模块基于文件信息和文件类型 2 种不同分类方式分类。将有敏感信息的文件靶向提取并单独存放，例如 Shell、Keyword、Admin、Passwd 等；将文件类型一致的文件批量提取并存入同类文件夹，如固件漏洞分析中使用较多的 bin 文件、conf 文件、cgi 文件等，从而实现操作系统层面的文件自动分类。

FTA 在对文件自动分类时输出敏感信息例如 Passwd、Key、Shell Scripts、Bin、IP Addresses、URL 等信息，以及包含这些信息的文件名称、路径等，将输出信息整合生成文本以便存储分析结果。

FTA 的自动分类模块能快速获取批量固件解析后的敏感信息以及目标文件，最终靶向提取关键目标。相较于重复搜索多类文件，批量靶向提取方法具有速度快、可视化、效率高的优势，大幅减少了解析文件所消耗的时间和磁盘空间，显著提高了批量固件分析的工作效率。

2.4 可扩展文件系统提取方法

2.4.1 固件识别技术

不同嵌入式设备的功能需求差异大，设备固件型号种类多，目前的解包工具不能识别不同固件所有种类的文件类型，在进行批量固件分析时会出现大量固件识别失败的情况，主要原因是在扫描分析过程中无法识别加密、混淆特征值的文件系统。本文基于文件系统和压缩文件识别原

理，研究 FTA 靶向识别技术，实现可扩展文件系统提取。

提取文件系统的原理是基于特征字段匹配。FTA 通过构造可扩展的 Magic 字段特征库，对固件二进制文件初始位置的特征码进行匹配，同时找到根文件系统起始地址以及地址范围，并将起始地址整段数据提取并保存，从而获得二进制镜像文件内的根文件系统（Magic 字段又称特征值，指文件最开头几个用于唯一区别其他文件类型的字节，可以区别不同的文件）。具体流程如图 5 所示。

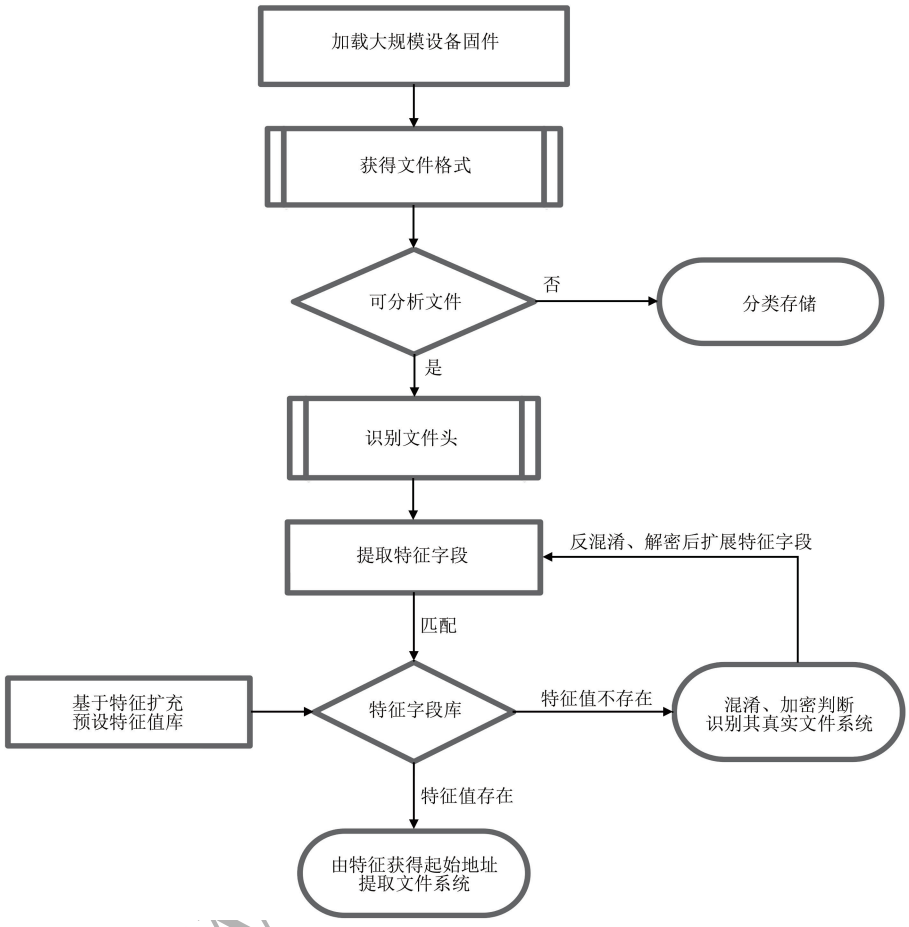

图 5 可扩展文件系统提取方法流程

Fig. 5 Extensible file system extraction method flow

识别批量固件的文件系统时，先获取文件格式并初步判断，保留 bin、hex 等可分析文件，进入下一步；自动分类存储不可分析文件，例如 txt、docx 等。

对可分析文件，获取文件头二进制信息，并提取特征字段，与预设字段库中的 Magic 值遍历匹配。特征字段库由 FTA 解析系统预设存储，包含 Squashfs、JSSF2 等主流文件系统和部分自制文件系统的特征字段。通过特征值字段库匹配后，有 2 种结果：

1）对匹配成功的文件，即特征值存在的文件，调用不同的文件识别插件进行靶向提取，找到根文件系统的起始地址及范围，靶向提取出文件系统并进行后续分析。

2）对匹配失败的文件，即特征值不存在的文件，初步判定为以下 3 种类别：

（1）通过特征值混淆后的主流文件系统，如对 Squashfs 的首部特征值“sqlz”混淆为“zlqs”，即文件系统的起始地址和大小端类型等格式都与“sqlz”型 Squashfs 相同，仅对 Magic 值混淆导致主流解析工具识别失败。该情况下 FTA 通过靶向识别插件提取混淆 Magic 值的文件系统，并反馈新的 Magic 值及其文件系统类型至特征字段库中，实现强化反馈，不断完善特征值库。

（2）对于加密后的文件，识别解析方法在第 3 章基于邻近版本的固件靶向解密中实现。

（3）对自制文件系统或冷门文件系统，初始预设特征字段库中没有相关信息，FTA 记录并提醒分析人员手动识别并扩充特征值字段库。

基于文件系统靶向识别方法，为了对更多混淆文件进行识别，提高批量固件解析的成功率和特征值匹配效果，设计实现 2 种靶向识别插件，用

于对 tar 文件和 Squashfs 文件系统进行靶向识别，基于特征值进行多种进制转化从而实现特征值匹配，对 Magic 值转换识别的原理如图 6 所示。

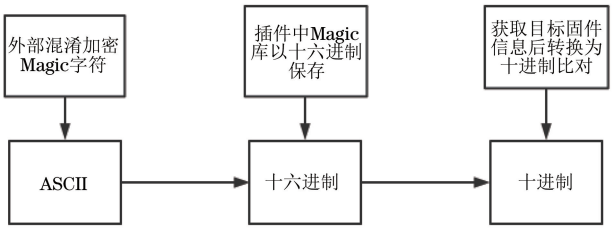

图 6 Magic 值转换识别原理

Fig. 6 Magic value conversion recognition principle

2.4.2 对混淆特征值的 tar 文件靶向识别方法

为增强对 tar 类型文件的靶向识别能力，研究压缩算法及签名和 tar 的特征值描述信息，设计 FTA 系统针对 tar 文件进行靶向识别，常见的压缩算法及签名如表 4 所示。

表 4 常见压缩算法及签名

Table 4 Common compression algorithms and signatures

压缩算法	偏移	类型	Magic 字段
7-zip	0	String	7z\274\257\047
Lzip	0	String	LZIP
LZMA	0	String	\xFFLZMA\x00
tar	0	String	ustar
JAR	0	Belong	0xcafed00d
Zilb	0	Beshort	0x789C,0x78DA

tar 文件是一种将多个文件和目录打包成一个单一的归档文件的文件格式，常见于 Unix 和 Linux 系统中，存在许多敏感信息，因此对其进行靶向识别非常重要。

固件中压缩了许多 tar 文件，一旦混淆其头部特征值，固件解包工具就无法识别并解包，导致大量被打包于 tar 文件中的敏感文件无法被提取。本节主要针对 tar 文件分析研究，实现靶向识别。

tar 文件的 Magic 值为“ustar”，并在文件描述中以“posix tar archive”字符串开头，Binwalk 识别解包时仅对文件描述进行判断，匹配则进行文件分块解析，否则不识别；因此研究基于扩展可识别特征值库方法，设计 FTA 自动解析系统的靶向识别模块，对 tar 文件的 Magic 值进行扩充，并在 FTA 识别过程中分析文件描述信息和 Magic 值，从而完整地识别并解包混淆 tar 文件。

FTA 靶向识别 tar 文件时，读取前 512 位信息，首先判断文件描述是否匹配“posix tar archive”，若匹配则视为 tar 文件，随后读取 tar 文件头结构、获取文件的大小转换为块数，更新下一个文件偏移量循环识别，并在缓冲区中搜索敏感字符串进行后续分析；若不匹配则比对缓冲区中 tar 文件特征值对应位置处的 Magic 值是否在预设的特征值库中，通过预设特征值库进行反混淆扩充以实现靶向识别 tar 文件。FTA 系统在靶向识别中扩充了 vtubs、abcde 等特征值，从而强化靶向识别能力。

2.4.3 对混淆特征值的 Squashfs 文件靶向识别方法

为增强 FTA 对 Squashfs 的靶向识别能力，研究 Squashfs 的二进制文件结构，Squashfs 的头部偏移信息如表 5 所示。

表 5 Squashfs 结构

Table 5 Squashfs structure

偏移量	类型	值	含义
0	String	sqsh	Squashfs file system; big endian
0	String	hsqs	Squashfs file system; little endian
0	String	sqlz	Squashfs file system; big endian; lzma compression
0	String	qshs	Squashfs file system; big endian; lzma signature
0	String	tsqh	Squashfs file system; big endian; DD-WRT signature
0	String	hsqt	Squashfs file system; little endian; DD-WRT signature
0	String	shsq	Squashfs file system; little endian; non-standard signature
>28	beshort	x	version %d
>28	beshort	>3	compression:
>>20	beshort	1	\bgzip
>>20	beshort	2	\blzma

分析表明，Squashfs 文件结构的特征值种类较多，有标准大端 sqsh、标准小端 hsqs、LZMA 压缩的大端 qshs、sqlz、小端 shsq、DD-WRT 固件小端 hsqt、DD-WRT 固件大端 tqsh 7 种，其他偏移量如

版本号、压缩方式等信息位置比较固定，因此同样可以基于扩展可识别特征值库方法完成对混淆 Squashfs 的靶向识别。

FTA 系统基于 Squashfs 结构信息，预设特征值库，记录特征值及相应偏移量对应的含义。在扫描文件后，通过 FTA 预设库中偏移量的值以及含义遍历检索待识别文件在相应偏移量处的 Magic 值并匹配，比对一致则依据 Squashfs 结构信息对待分析文件完成靶向识别；特征值匹配不一致时继续对后续偏移量的值与预设信息库进行比对，若仅特征值不存在，而其他偏移量均匹配符合，则靶向识别为混淆后的 Squashfs，同时记录其前 4 Byte 作为新的特征值来反馈扩充预设特征值库；如果其他偏移量均匹配失败，则判定为非靶向识别的文件系统。

3 固件靶向解密方法

3.1 固件库构建

在批量固件解析中存在大量加密固件无法正确解析。为实现固件解密，构建版本连续的固件库，通过 3 种途径获取固件，如图 7 所示。

1）爬虫获取：通过设计网络爬虫脚本自动从各设备厂商的固件发布页面获取固件的初始版本和升级版本等信息。

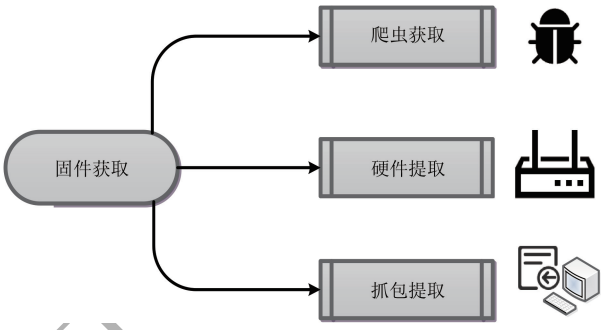

图 7 固件库构建方法

Fig.7 Firmware library building methodology

2）硬件提取：通过手动从开发板中提取固件，利用固件系统中的 UART、SPI、JTAG 接口来获取固件。

3）抓包提取：利用 WireShark 软件配合热点抓取设备固件升级的更新包。

通过以上方式从网络和设备中获取大量固件，通过 FTA 自动化固件解析工具对固件进行批量解析。利用浅度、深度提取功能提取信息后进行切分归类处理。对数据库所需的固件信息定向收集并保存至数据库，完成固件库的构建。固件库的基本结构如图 8 所示。

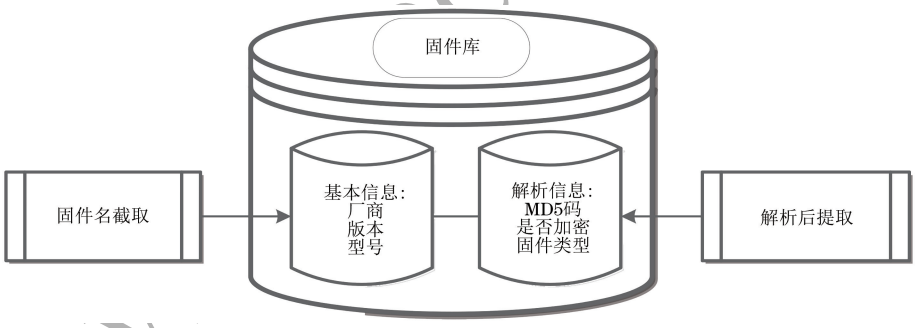

图 8 固件库基本结构

Fig.8 Firmware library basic structure

1）对所需的固件基本信息，例如固件厂商、型号、版本等通过固件名切分截取。使用靶向提取功能获取不同固件的说明文档，并在文档中定向搜索版本和型号的关键词。

2）对所需的固件解析信息，例如固件的特征值、MD5 码、是否加密、固件类型等（Ⅰ、Ⅱ、Ⅲ型固件），在自动分类和信息输出的文档中进行提取。

3.2 基于邻近版本的固件靶向解密方法

针对固件加密导致固件解析工具无法识别或者识别出错，导致难以进行后续代码分析和漏洞检测的问题，研究固件加密机制及靶向解密方法。

对固件库中的固件分析发现，固件加密一般使用 AES 或 DES 等对称加密算法，保护固件中的敏感算法、配置文件等重要信息。加密固件的固件头、字符串信息、Magic 签名等二进制数据被修改，即使成功提取也只能看到密文数据。经过实验验证，当前主流固件分析工具无法直接识别加密固件的文件系统、压缩格式、版本等信息。

固件加密流程如图 9 所示，分为 2 种：

1）物联网设备固件在厂商设计出厂时并未加密，在初始版本中也没有任何解密程序。在用户需求逐渐增多的情况下，不断对固件打补丁升级并实现服务器云端更新，在某个较新的版本（过渡版本）中进行了固件升级，服务器云端上传发布后，内嵌解密程序供用户下载，以便将来对加密固件更新。由于过渡版本已经植入解密程序，厂商

后续发布的固件都是加密固件，没有解密程序则无法解析和提取文件系统。针对此类情况，针对过渡版本入手，从包含解密程序的未加密过渡版本中提取到解密程序，然后通过固件仿真来解密后续的新版本固件。

2）物联网设备固件在厂商设计出厂时已经加密，但是厂商认为加密方式不再安全可靠，因此在固件升级更新的过程中更改为新的加密方案，并发布一个未加密的过渡版本（例如图中的未加密 V2.0 版本），其中包含新的解密程序，且在后续更新中都发布加密版本实现安全加密更新。在这种情况下，寻找到过渡版本，通过对其固件解析提取解密程序，并应用到后续的新版本加密固件中，即可实现解密。

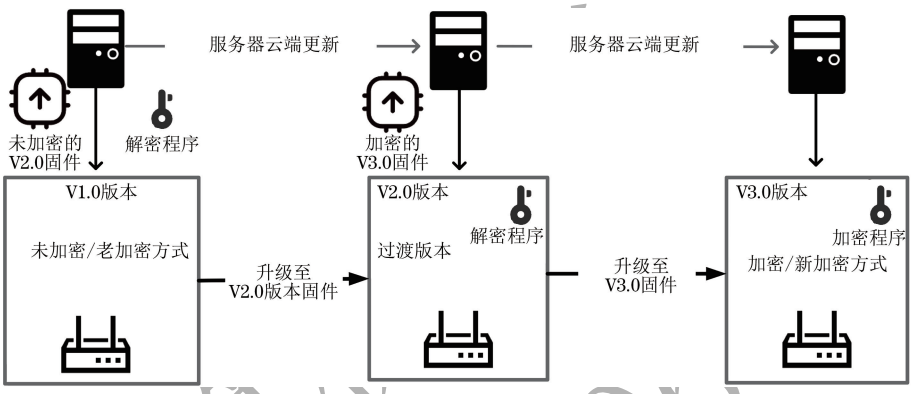

图 9 常用的固件加密流程

Fig. 9 Commonly used firmware encryption flow

对 2 种情况分析发现，过渡版本固件非常关键。基于固件库构建，搜集存储大量不同版本的固件，为本节固件解密方法实现提供数据支撑。寻找过渡版本通常依靠厂商更新固件后发布的相关公告获取信息，但这种方式效率低，执行难度大。为实现自动固件解密，FTA 系统设计邻近版本递归搜索的靶向解密方法，基于固件库支撑，自动检索不同版本的固件获取过渡版本。实现流程如图 10 所示。

以构建信息分类存储的固件库作为预处理阶段，根据固件名称、型号等信息分类存储在数据库中，作为靶向解密的基础研究。FTA 系统的固件靶向解密方法共分为 5 个阶段：

1）分析阶段：固件解析初始阶段，对固件熵分析来判定熵值大小，若为低熵则判定为非加密固件；若为高熵则判定为加密固件或压缩文件。

2）解析阶段：实现初步判断后，对固件靶向解析，若固件解包成功，则判定为非加密版本或过渡版本，归类为可解析固件，否则判断为加密固件。

3）搜索阶段：在数据库搜索邻近版本，直至找到能解析的固件为止。

4）递归阶段：搜索阶段后，对搜索到的不同版本固件递归分析，即回跳执行分析阶段，在解析阶段递归解析，对固件库中所有同型号不同版本固件进行批量解析，直至识别成功，记录其熵值，定义为低熵并停止递归。

5）解密阶段：对解包成功的过渡版本固件靶向提取关键信息，判断加密方式并检索密钥。由于固

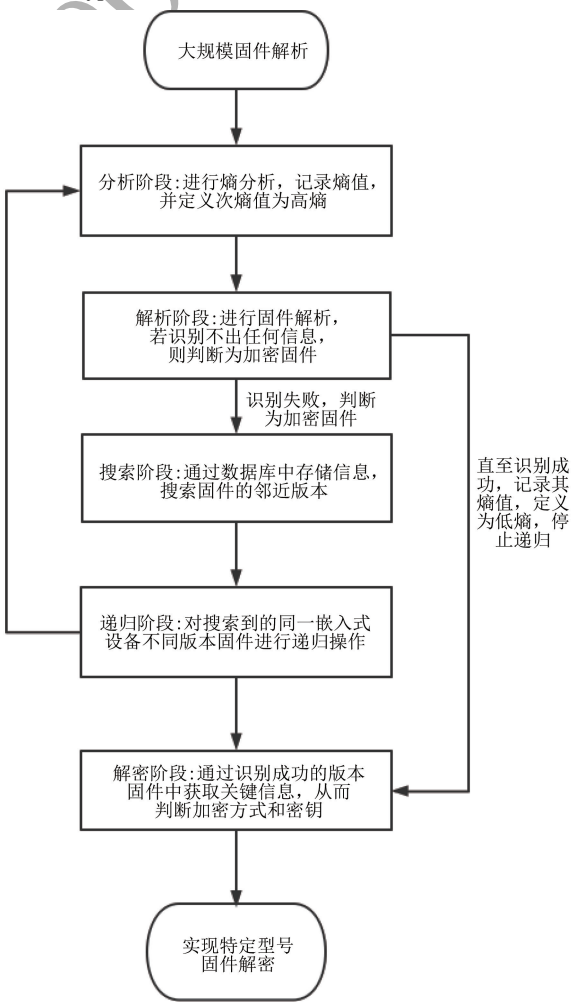

图 10 基于相邻版本的固件靶向解密方法

Fig. 10 Firmware targeted decryption method based on neighboring versions

件版本更替时有大量重叠的信息，因此只需检索任一可被 FTA 靶向识别的版本固件，分析其与相邻版本之间的差异和固件更新的内容，提取密钥并识别加密算法，最终完成特定型号固件解密。

4 实验结果与分析

4.1 实验设计和评估标准

针对当前嵌入式设备固件批量分析存在的效率低、过程复杂、自动化程度低等问题，利用本文所提技术展开实验。实验流程基于常规固件解包过程，主要对靶向提取模块、自动分类及关键信息输出、靶向识别模块、基于邻近版本的固件靶向解密模块和批量固件快速解析进行实验。

实验选取通过爬虫从供应商网站获得的固件以及从路由器等硬件设备手动提取的固件作为数据集，详细信息如下：

1)来自供应商网站的固件数据集。

(1)从 TP-Link 处共获取了 523 个固件二进制文件，其中包括各种路由器固件，例如 VPN 路由器、无线迷你路由器、商用路由器、家庭路由器、交换机和智能摄像头等。

(2)从 Tenda 网站收集 119 个固件。收集的数据集包括移动 Wi-Fi 热点、无线智能路由器、网络交换机、千兆桌面交换机和 IP 网络摄像头的固件。

(3)从 D-Link[31] 处下载 263 个路由器固件二进制文件，其中也包含加密的二进制文件。

(4)从 Netis、Mercury、Netgear、Belkin、飞鱼星和思科供应商网站共获取 164 个固件二进制文件，其中包括路由器、交换机、WiFi 信号放大器和网络适配器。

(5)从 NETGEAR 网站下载了 306 个固件文件，其中大部分是路由器的二进制镜像，同时也有压缩文件和二进制 bin 文件。

2)来自供应商实体设备的固件数据集。

从供应商网站爬取固件的同时，手动从实体硬件设备中提取不同型号设备的嵌入式固件。

(1)UART 串口提取 VxWorks 路由器固件。

(2)利用 J-Llink 从 JTAG 接口提取智能终端固件。

(3)通过飞线法以及编程器读取并识别智能摄像头的二进制固件。

(4)通过拆焊 Flash 芯片获得通信设备中的 VxWorks 固件。

由于 Binwalk 未设类似 FTA 的冗余清除机制，大规模固件解析会导致磁盘存储空间不足，为与 FTA 做对照测试，在 2 种方式获取的数据集中，选取包含不同的系统架构(MIPS、ARM 等)、操作系统内核(VxWorks、Linux 等)和各种文件系统(Squashfs、JFFS2、YAFFS 等)的各类型固件共 100 个作为实验组进行批量固件自动化解析实验。

实验评估指标：根据不同的实验流程和结果，选取靶向提取正确率、自动分类成功率、速度性能提升率作为 FTA 固件自动解析实验的评估指标。

在靶向提取性能评估中，统计 FTA 靶向提取的文件数量与实验组中目标文件的数量之比作为靶向提取的正确率。

在自动分类性能评估中，计算分类成功率作为实验评估指标，统计 FTA 自动分类的各文件类型数量与实验组中该类型文件的个数之比作为自动分类成功率。

在批量固件快速解析性能评估中，记录浅度解析和深度解析实验组固件的用时，以 FTA 优化时间与 Binwalk 用时之比作为速度性能提升率。

4.2 功能对比评估

基于面向异构固件的高效靶向分析技术，设计实现 FTA 自动化固件解析系统，表 6 为 FTA 和 Binwalk 实现功能上的对比。

表 6 FTA 与 Binwalk 功能对比

Table 6 Comparison of FTA and Binwalk functionalities

功能	Binwalk	FTA
固件解析	手动输入	多粒度固件解析
解压方式	特征识别	基于特征识别的浅度和深度解压
解析信息处理	无分类输出	自动输出敏感信息识别存储
信息存储	无	根据文件名称信息和文件内容信息构建固件库
文件检测	无	依据白/黑名单快速分类
特征值库	预设无法扩充	基于解析反馈的可扩充特征值库
固件中文件识别	不支持	基于特征值库靶向识别 Squashfs、tar 等文件并扩充 9 类混淆文件识别
加密识别	无	基于邻近版本的固件解密
固件中文件输出	固件完全解析后全部输出	目标文件靶向提取输出
固件解析规模	单一固件解析	批量固件解析
固件平均解析速度/(个·min^{-1})	23	39(提升 1.7 倍)

从表中可以看出，由于 FTA 自动解析工具基于对固件解包原理深入研究实现多粒度固件分析和靶向识别方法，集成了信息存储、自动分类、加密识别等多个功能模块，较传统工具 Binwalk 的功能更多样化，便于分析人员进行固件批量分析处理，并简化了操作流程。在解析性能上，强化了对混淆固件的识别解析能力和对加密固件的一定程度上的解密提取功能，并且对信息和文件输出进行优化，最终获得相对完整详细的敏感信息报告。

4.3 靶向提取性能评估

FTA 的重要功能之一是靶向提取，在固件分析流程中完成批量自动解析并检索和提取解包后存在的目标文件。实验选取漏洞分析中脆弱性较高的文件类型和名称作为关键词，FTA 对实验组批量解析并根据不同的关键词分组进行 4 组靶向提取实验，根据实验评估指标计算正确率，评估 FTA 靶向提取性能。

靶向提取模块测试结果如图 11 所示。

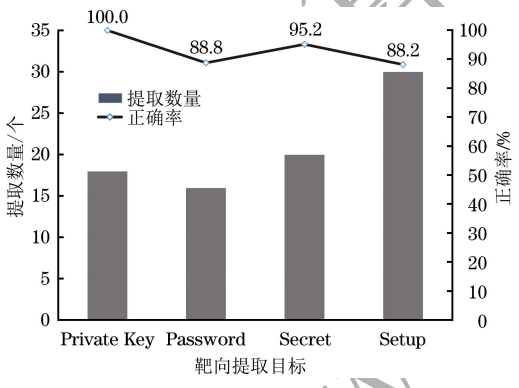

图 11 靶向提取模块测试结果

Fig. 11 Targeted extraction module test results

从图中可以直观看出 FTA 的靶向提取模块对 Password、Secert、Setup、Private Key 4 组关键词的目标的成功率均在85%以上，提取错误的文件较少，其中准确率较低的是关键词为 Password 的实验组。

提取失败的原因主要是部分固件解包不全以及靶向提取目标文件时存在文件名识别不够精确导致遗漏。主流解析工具 Binwalk 并不能有针对性地批量提取文件，在漏洞分析时效率不高。实验结果表明 FTA 可以帮助分析人员实现对关键词文件的靶向提取工作，从而提高固件分析的效率。

4.4 自动分类及关键信息输出性能评估

设计自动分类及关键信息输出性能评估实验，测试 FTA 对解包文件自动分类及输出的情况，评估指标为分类成功率，测试结果如图 12 所示。

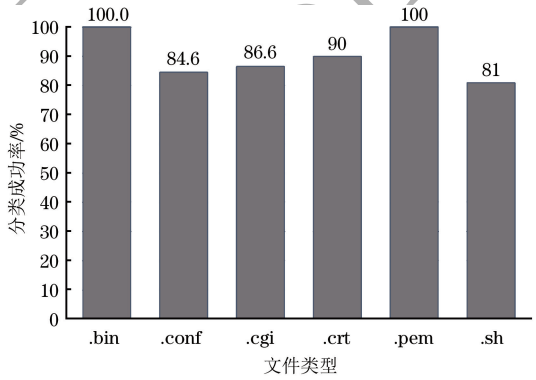

图 12 文件自动分类模块测试结果

Fig. 12 Automatic document categorization module test results

实验选取 bin、conf、cgi、crt、pem 5 类敏感文件类型作为测试。实验结果表明 FTA 对批量固件解析提取后能准确对选取的 5 类文件进行正确的分类，遗漏文件较少。

经分析，分类测试遗漏文件主要原因是部分文件提取权限不够以及存在同名文件导致遗漏。

输出模块能根据固件提取到的文件进行关键信息输出，依据文件类型和关键信息分类输出到 json 文件中，实验中输出的分类和描述如表 7 所示。

表 7 关键信息输出性能评估

Table 7 Key message output performance evaluation

输出内容	描述
SSL related files	协议和加密通信相关的文件，在建立和维护安全的 SSL 连接时起着重要的作用，是比较重要的分析内容
Database related files	数据库相关文件，提取后用于分析数据库中的敏感信息
Shell scripts	用于自动化执行一系列 Shell 命令的脚本文件，分析固件中的 Shell 脚本可以获取和系统配置和初始化有关的信息，有助于安全漏洞和脆弱性检测
Passwd	固件中和密码有关的文件信息
Admin/root	解析后与权限有关的文件类型
Ip addresses	解析后获取的有用 IP 地址
URLs	提取文件中包含的 URL 链接
Emails	提取后获得的有关邮箱，可以通过社会工程学进行攻击

基于关键词自动遍历检索原理，FTA 能提取并靶向输出分类后的文件，然后汇总以便于分析人员查找浏览所需的信息和信息所在的文件地址。

由于 Binwalk 只能按照原固件加载时的存储结构将文件全部提取出来，缺少文件处理模块，因此无法对提取后大量文件分类输出。FTA 自动解析工具能在批量固件解析后按文件名和文件类型将高脆弱性的文件分类存储，帮助分析人员快速寻找所需的文件类型。

4.5 识别功能对比评估

针对固件特征值混淆识别进行实验，基于预设 Magic 库中包含的 Squashfs 大小端、压缩方式等信息，对混淆后的特征值根据偏移量的值与含义匹配识别并提取相应的文件系统。FTA 与 Binwalk 的识别能力对比结果如表 8 所示。

表 8 识别能力对比

Table 8 Comparison of recognition ability

文件类型	原特征值	混淆特征值	FTA	Binwalk
Squashfs	sqlz	zlqs	识别正确；big endian；lzma compression	无法识别
	shsq	shsa	识别正确；little endian；non-standard signature	无法识别
	hsqs	hsqq	识别正确；little endian	无法识别
	sqsh	ssqs	识别正确；big endian	无法识别
	qshs	hhsq	识别正确；big endian；lzma signature	无法识别
	tsqh	ssqt	识别正确；big endian；DD-WRT signature	无法识别
	hsqt	hsqz	识别正确；little endian；DD-WRT signature	无法识别
tar	ustar	tarus	识别正确	无法识别
	ustar	vstub	识别正确	无法识别

实验结果表明：对 Squashfs 特征值进行混淆后，Binwalk 对 zlqs、shsa 等非标准 Magic 值出现了无法识别的情况，原因在于 Binwalk 没有对 Squashfs 进行特征值扩充，特征值库中并没有相应的匹配内容，导致识别失败，不能正确解析出 Squashfs；对于 tar 文件，混淆特征值后，Binwalk 均无法识别成功，这是因为 Binwalk 对压缩文件的识别基于签名匹配，调用相应的算法进行解压，并没有预设特征库。

由于 FTA 系统通过针对 tar 文件和 Squashfs 预设特征库进行偏移内容匹配，因此对混淆后的特征值例如 tarus 和 vstub 能够正确识别并进行后续解析，完成靶向识别功能，与 Binwalk 的解析情况对比有明显的提升。

4.6 靶向解密性能评估

对 FTA 进行固件解密性能评估中，选取数据集中具有代表性的不同厂商加密固件作为实验组，通过在固件库中自动搜寻加密固件的相邻版本并递归解析检索过渡版本，分析过渡版本并提取解密程序，完成对测试集中的固件样例进行解密。

以 D-Link 品牌固件为例进行测试分析，表 9 是选取 D-Link 的不同型号固件进行加解密性能测试的实验结果，实验中随机选取不同的加密版本，并基于固件库中的邻近版本进行递归搜寻过渡版本，在找到过渡版本后通过提取解密程序 imgdecrypt、获取 key 来解密加密版本的固件。

表 9 靶向解密性能评估结果

Table 9 Targeted decryption performance evaluation results

固件型号	加密版本	未加密版本(过渡)	解密结果
DIR-3040	v1.13B03	v1.13B02	解密成功
DIR-822	v3.15B02	v303WWb04	解密成功
DIR-878	FW120B05	FW104B05	解密成功
DIR-882	v1.20B06	v1.10B02	解密成功
EA6100	v1.1.6.181939	v1.1.5.172244	解密成功

实验结果表明：对于固件库中的主流厂商固件版本型号，FTA 系统对 DIR 系列固件解密成功率很高，主要原因是固件库中对该类型号固件搜集较全面，FTA 系统基于邻近版本的固件靶向解密方法对固件解包时能获取过渡版本的解密程序，实现成功解密，在选取的 20 个不同厂商固件进行解密的实验测试中的成功率为 80%；对于解密失败的固件，分析原因是其固件版本较少，固件获取困难，导致无法匹配过渡版本，因此无法提取解密程序实现解密。

4.7 批量固件快速解析性能评估

为评估 FTA 和 Binwalk 对批量固件进行解析的速度性能，对实验组中 100 个固件，测试 FTA 自动化固件解析工具的浅度解析和深度解析模块的解析速度性能，分别与 Binwalk 普通文件提取和递归扫描提取的速度进行对比。测试结果如图 13 所示。

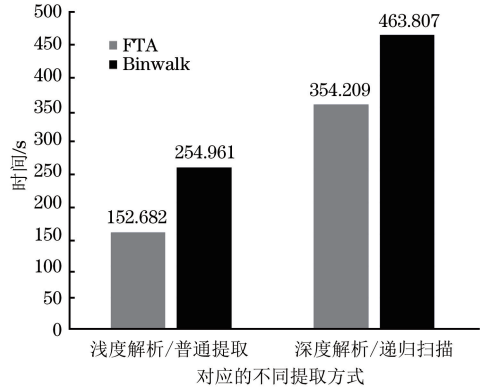

图 13 批量固件解析速度

Fig. 13 The speed of batch firmware parsing

测试结果表明，在以 100 个文件为实验组进行批量固件解析的 2 组对照组中，FTA 对同一组文件的提取用时均大幅少于 Binwalk，下面对 2 组结果进行分析：

第 1 组：FTA 的浅度解析模块和 Binwalk 的普通提取功能的性能测试。结果表明，Binwalk 对文件进行提取的用时为 254.961 s，而 FTA 对文件浅度初步提取的用时为 152.682 s，减少用时 102.279 s，速度性能提升 40.15%，同时能提取固件文件内核并保存。主要原因是 FTA 预设黑名单和白名单，在浅度解析过滤杂项文件的同时清除冗余文件，因此在批量固件的提取中会有比较明显的速度优势。

第 2 组：FTA 的深度解析模块和 Binwalk 的递归扫描功能的性能测试。结果同样表明，FTA 解析用时 354.209 s，大幅少于 Binwalk 递归扫描用时(463.807 s)，共减少 208.846 s，性能提升 45.03%，同时也能提取出文件中的各种信息，例如文件系统、符号表、U-BOOT 等。主要原因是 FTA 在深度解析时对嵌套文件识别，将判断非压缩类型的文件转入浅度匹配解析，从而减少了用时，提高了解析效率。

以上 2 组对照测试的结果表明 FTA 在设置黑名单和白名单、关键字匹配，并优化解析流程后，在浅度解析和深度解析上的速度都明显优于 Binwalk，综合速度性能提升 42.59%，大幅提升了批量解析速度。

总体来说，FTA 在靶向提取、文件自动分类、关键信息输出、靶向识别、解密固件和批量固件提取等方面的性能测试中均表现良好，同时，通过与 Binwalk 在不同功能上对比结果来看，FTA 在批量固件解析、文件信息处理、识别内容扩展和敏感文件提取等方面均存在明显的优势。因此 FTA 系统能在大规模固件分析和漏洞挖掘中起到简化分析人员操作流程和大幅度提高效率的作用。

5 结束语

本文从批量固件安全分析出发，针对目前国内外前沿研究中大规模固件分析过程样本需求量大、人工检索目标文件效率低等问题，分析了固件获取及存储相关技术；针对已有工具 Binwalk 对文件系统支持数量少、解析效果不理想等不足，深入分析了固件解析技术原理、文件系统结构原理，基于这些原理，提出一种面向异构固件的高效靶向分析技术，设计实现 FTA 固件分析系统，实现靶向提取、文件自动分类、关键信息输出、靶向识别等功能；设计邻近版本的固件靶向解密方法，基于固件库支撑完成对特定版本固件的解密，旨在提高批量固件解析识别发现脆弱点的效率。

同时，本文的研究也有一些不足：由于搜集到的固件来自不同厂商，其采用的加密算法非标准且不透明，导致部分固件解析时出现异常，难以分析加密后的特征值，因此对加密固件识别扩展有限。下一步研究方向在于对 VxWorks 固件的加密机制和符号表恢复进行更深入的研究，增强 FTA 系统对 VxWorks 固件的靶向解析。

参考文献

[1] 中国通信标准化协会. 物联网操作系统安全白皮书[EB/OL]. (2022-09-19)[2024-03-11]. https://blog.nsfocus.net/wp-content/uploads/2022/09/iot-whitepaper.pdf.
China Communications Standards Association. IoT operating system security white paper[EB/OL]. (2022-09-19)[2024-03-11]. https://blog.nsfocus.net/wp-content/uploads/2022/09/iot-whitepaper.pdf. (in Chinese)

[2] MUENCH M, STIJOHANN J, KARGL F, et al. What you corrupt is not what you crash: challenges in fuzzing embedded devices[C]// Proceedings of the Network and Distributed System Security Symposium. Washington D. C., USA: IEEE Press, 2018: 1-10.

[3] 周诚远, 张慧翔, 李晓辉, 等. 浅析物联网设备固件保护与漏洞分析技术[J]. 保密科学技术, 2022(2): 45-50.
ZHOU C Y, ZHANG H X, LI X H, et al. Analysis on firmware protection and vulnerability analysis technology of Internet of Things equipment[J]. Secrecy Science and Technology, 2022(2): 45-50. (in Chinese)

[4] ANDREI C, JONAS Z, AURELIEN F, et al. A large-scale analysis of the security of embedded firmwares[C]// Proceedings of the 23th USENIX Security Symposium. Berkeley, USA: USENIX, 2014: 95-110.

[5] POOJA G, PRIYANKA G, GAYATRI C, et al. Steganography using bin-walk tool & its overview[J]. International Journal of Computer Science and Mobile Computing, 2021, 10(6): 90-96.

[6] DAVID B, IVAN J, THANASSIS A, et al. BAP: a binary analysis platform[C]// Proceedings of CAV 2011. Berlin, Germany: Springer, 2011: 463-469.

[7] HEITMAN C, ARCE I. BARF: a multiplatform open source

binary analysis and reverse engineering framework[EB/OL]. (2021-05-26)[2024-03-11]. https://raw.githubusercontent.com/programa-stic/barf-project/master/doc/papers/barf.pdf.
[8] HAIDER A, MUHAMMAD S, MALIHA S, et al. DUDE: decryption, unpacking, deobfuscation, and endian conversion framework for embedded devices firmware [J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(4): 99-101.
[9] REDINI N, MACHIRY A, WANG R Y. Karonte: detecting insecure multi-binary interactions in embedded firmware[C]// Proceedings of 2020 IEEE Symposium on Security and Privacy. San Francisco, USA: IEEE Press, 2020: 1544-1561.
[10] Mitre. Cve-2019-5137 [EB/OL]. (2020-03-22)[2024-03-11]. https://nvd.nist.gov/vuln/detail/CVE-2019-5137.
[11] TIEN C W, TSAI T T, CHEN I Y, et al. UFO-hidden backdoor discovery and security verification in IoT device firmware [C] // Proceedings of the IEEE International Symposium on Software Reliability Engineering. Memphis, USA: IEEE Press, 2018: 18-23.
[12] SHWARTZ O, MATHOV Y, BOHADANA M, et al. Reverse engineering iot devices: effective techniques and methods[J]. IEEE Internet of Things Journal, 2018, 5(6): 4965-4976.
[13] GAMBLIN J. Mirai-Source-Code: leaked mirai source code for research/ioc development purposes[EB/OL]. (2020-02-27)[2024-03-11]. https://github.com/jgamblin/Mirai-Source-Code.
[14] COSTIN A, ZARRAS A, FRANCILLON A. Automated dynamic firmware analysis at scale: a case study on embedded web interfaces[C]// Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security. New York, USA: ACM Press, 2016: 437-448.
[15] ELSABAGH M, JOHNSON R, STAVROU A, et al. FIRMSCOPE: automatic uncovering of privilege-escalation vulnerabilities in pre-installed apps in android firmware[C]// Proceedings of the 29th USENIX Security Symposium. [S.l.]: USENIX, 2020: 1-10.
[16] OBERMAIER J, HUTLE M. Analyzing the security and privacy of cloud-based video surveillance systems[C]//Proceedings of the 2nd ACM International Workshop on IoT Privacy, Trust, and Security. New York, USA: ACM Press, 2016: 22-28.
[17] HEMEL A. Binary analysisng [EB/OL]. (2021-05-25)[2024-03-11]. https://github.com/armijnhemel/binaryanalysisng.
[18] DESIGN E. Cryptographic techniques for safer firmware [EB/OL]. (2023-03-05)[2024-03-11]. https://www.electronicdesign.com/technologies/embedded-revolution/article/21163055/neuronicworks-cryptographic-techniques-for-safer-firmware.
[19] 赵亚新, 郭玉东, 舒辉. 基于 JTAG 的嵌入式设备固件分析技术[J]. 计算机工程与设计, 2014, 35(10): 3410-3415.
ZHAO Y X, GUO Y D, SHU H. Analysis technology of embedded device firmware based on JTAG[J]. Computer Engineering and Design, 2014, 35(10): 3410-3415. (in Chinese)
[20] PARTNERS P T. How to do firmware analysis. tools, tips, and tricks[EB/OL]. (2021-05-25)[2024-03-11]. https://www.pentestpartners.com/security-blog/how-to-do-firmware-analysis-tools-tips-and-tricks/.
[21] 于颖超, 陈左宁, 甘水滔, 等. 嵌入式设备固件安全分析技术研究[J]. 计算机学报, 2021, 44(5): 859-881.
YU Y C, CHEN Z N, GAN S T, et al. Research on the technologies of security analysis technologies on the embedded device firmware [J]. Chinese Journal of Computers, 2021, 44(5): 859-881. (in Chinese)
[22] CHEN D D, EGELE M, WOO M, et al. Towards automated dynamic analysis for linux-based embedded firmware[C]//Proceedings of 2016 Network and Distributed System Security Symposium. San Diego, USA: Internet Society, 2016: 1-10.
[23] KIM M, KIM D, KIM E, et al. FirmAE: towards large-scale emulation of IoT firmware for dynamic analysis[C]// Proceedings of the Annual Computer Security Applications Conference. New York, USA: ACM Press, 2020: 733-745.
[24] 朱晓东. 二进制代码相似性分析关键问题研究[D]. 郑州: 战略支援部队信息工程大学, 2021.
ZHU X D. Research on key problems of binary code similarity analysis[D]. Zhengzhou: Information Engineering University, 2021. (in Chinese)
[25] 于颖超, 甘水滔, 邱俊洋, 等. 二进制代码相似度分析及在嵌入式设备固件漏洞搜索中的应用[J]. 软件学报, 2022, 33(11): 4137-4172.
YU Y C, GAN S T, QIU J Y, et al. Binary code similarity analysis and its applications on embedded device firmware vulnerability search[J]. Journal of Software, 2022, 33(11): 4137-4172. (in Chinese)
[26] PEWNY J, GARMANY B, GAWLIK R, et al. Cross-architecture bug search in binary executables [C] // Proceedings of the IEEE Symposium on Security and Privacy. San Jose, USA: IEEE Press, 2015: 709-724.
[27] PECKOL J K. Embedded systems: a contemporary design tool[M]. [S. l.]: John Wiley & Sons, Inc., 2019.
[28] 朱晓东, 尹青, 常瑞, 等. 基于结构化特征库的递进式固件格式解析[J]. 武汉大学学报(理学版), 2017, 63(2): 125-132.
ZHU X D, YIN Q, CHANG R, et al. Structured feature library-based progressive firmware format parsing [J]. Journal of Wuhan University (Natural Science Edition), 2017, 63(2): 125-132. (in Chinese)
[29] CYR B, MAHMOD J, GUIN U. Low-cost and secure firmware obfuscation method for protecting electronic systems from cloning[J]. IEEE Internet of Things Journal, 2019, 6(2): 3700-3711.
[30] ARMIJN H. Introducing the binary analysis tool[EB/OL]. (2013-05-27)[2024-03-11]. http://events.static.linuxfound.org/sites/events/files/als13_hemel_bat.pdf.
[31] D-link. D-link series dir-867[EB/OL]. (2022-03-27)[2024-03-11]. https://support.dlink.com/productinfo.aspx?m=DIR-882-US.

编辑 金胡考