



网络与信息安全前沿技术丛书

网络协议 逆向分析及应用

吴礼发 洪征 潘璠 著

Network Protocol
Reverse Analysis and Application



国防工业出版社
National Defense Industry Press



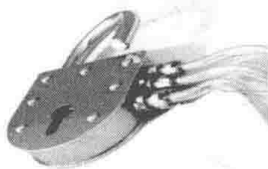
国防科技图书出版基金

网络与信息安全前沿技术丛书

吴礼发 洪 征 潘 璠 著

网络协议逆向 分析及应用

Network Protocol Reverse Analysis and Application



网络监听、网络对抗、网络管理、恶意代码分析、软件安全漏洞挖掘等应用领域常常需要分析未知协议的格式，高效、准确的自动化协议逆向分析技术一直是人们追求的目标。本书是国内第一本全面介绍自动化网络协议逆向分析及应用的学术专著，反映了协议逆向分析领域的最新研究成果，可作为从事软件及协议逆向分析、网络管理、网络安全与对抗等方向的教学、科研及工程技术人员的参考书。



国防工业出版社
National Defense Industry Press

· 北京 ·

图书在版编目(CIP)数据

网络协议逆向分析及应用 / 吴礼发,洪征,潘瑶著.
—北京:国防工业出版社,2016.1
(网络与信息安全前沿技术丛书)
ISBN 978-7-118-10574-2

I. ①网... II. ①吴... ②洪... ③潘... III. ①计算机
网络-通信协议 IV. ①TP915.04

中国版本图书馆 CIP 数据核字(2015)第 312805 号

※

国防工业出版社出版发行
(北京市海淀区紫竹院南路 23 号 邮政编码 100048)
北京嘉恒彩色印刷有限责任公司
新华书店经售

*

开本 710×1000 1/16 印张 20 $\frac{3}{4}$ 字数 388 千字
2016 年 1 月第 1 版第 1 次印刷 印数 1—3000 册 定价 99.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)88540777

发行邮购:(010)88540776

发行传真:(010)88540755

发行业务:(010)88540717

致 读 者

本书由国防科技图书出版基金资助出版。

国防科技图书出版工作是国防科技事业的一个重要方面。优秀的国防科技图书既是国防科技成果的一部分,又是国防科技水平的重要标志。为了促进国防科技和武器装备建设事业的发展,加强社会主义物质文明和精神文明建设,培养优秀科技人才,确保国防科技优秀图书的出版,原国防科工委于1988年年初决定每年拨出专款,设立国防科技图书出版基金,成立评审委员会,扶持、审定出版国防科技优秀图书。

国防科技图书出版基金资助的对象是:

1. 在国防科学技术领域中,学术水平高,内容有创见,在学科上居领先地位的基础科学理论图书;在工程技术理论方面有突破的应用科学专著。
2. 学术思想新颖,内容具体、实用,对国防科技和武器装备发展具有较大推动作用的专著;密切结合国防现代化和武器装备现代化需要的高新技术内容的专著。
3. 有重要发展前景和有重大开拓使用价值,密切结合国防现代化和武器装备现代化需要的新工艺、新材料内容的专著。
4. 填补目前我国科技领域空白并具有军事应用前景的薄弱学科和边缘学科的科技图书。

国防科技图书出版基金评审委员会在总装备部的领导下开展工作,负责掌握出版基金的使用方向,评审受理的图书选题,决定资助的图书选题和资助金额,以及决定中断或取消资助等。经评审给予资助的图书,由总装备部国防工业出版社列选出版。

国防科技事业已经取得了举世瞩目的成就。国防科技图书承担着记载和弘扬这些成就,积累和传播科技知识的使命。在改革开放的新形势下,原国防科工委率先设立出版基金,扶持出版科技图书,这是一项具有深远意义的创举。此举势必促使国防科技图书的出版随着国防科技事业的发展更加兴旺。

设立出版基金是一件新生事物,是对出版工作的一项改革。因而,评审工作需

要不断地摸索、认真地总结和及时地改进,这样,才能使有限的基金发挥出巨大的效能。评审工作更需要国防科技和武器装备建设战线广大科技工作者、专家、教授,以及社会各界朋友的热情支持。

让我们携起手来,为祖国昌盛、科技腾飞、出版繁荣而共同奋斗!

国防科技图书出版基金
评审委员会

国防科技图书出版基金 第七届评审委员会组成人员

主任委员	潘银喜			
副主任委员	吴有生	傅兴男	杨崇新	
秘书长	杨崇新			
副秘书长	邢海鹰	谢晓阳		
委员	才鸿年	马伟明	王小谟	王群书
(按姓氏笔画排序)	甘茂治	甘晓华	卢秉恒	巩水利
	刘泽金	孙秀冬	芮筱亭	李言荣
	李德仁	李德毅	杨伟	肖志力
	吴宏鑫	张文栋	张信威	陆军
	陈良惠	房建成	赵万生	赵凤起
	郭云飞	唐志共	陶西平	韩祖南
	傅惠民	魏炳波		

《网络与信息安全前沿技术丛书》编委会

主 任 何德全

副主任 吴世忠 黄月江 祝世雄

秘 书 张文政 王晓光

编 委 (排名不分先后)

郭云飞	邢海鹰	胡昌振	王清贤	荆继武
李建华	王小云	徐茂智	吴文玲	郝 平
孙 琦	张文政	陈克非	杨 波	胡予濮
卿 昱	杨 新	肖国镇	陈晓桦	饶志宏
谢上明	周安民	许春香	唐小虎	曾 兵
曹云飞	陈 晖	周 宇	安红章	陈周国
王宏霞	霍家佳	董新锋	赵 伟	郑 东
郝 尧	李 新	冷 冰	穆道光	申 兵
汤殿华	张李军	胡建勇		

网络的触角正伸向全球各个角落,高速发展的信息技术已渗透到各行各业,不仅推动了产业革命、军事革命,还深刻改变着人们的工作、学习和生活方式。然而,在人们享受信息技术带来巨大利益的同时,一次又一次网络信息安全领域发生的重大事件告诫人们,网络与信息安全已直接关系到国家安全和社
会稳定,成为我们面临的新的综合性挑战,没有过硬的技术,没有一支高水平的人才队伍,就不可能在未来国际博弈中赢得主动权。

网络与信息安全是一门跨多个领域的综合性学科,涉及计算机科学、网络技术、通信技术、密码技术、信息安全技术、应用数学、数论、信息论等。“道高一尺、魔高一丈”,网络与信息安全技术在博弈中快速发展,出版一套覆盖面较全、反映网络与信息安全方面新知识、新技术、新发展的丛书有着十分迫切的现实需求。

适逢此时,欣闻由我国网络与信息安全领域著名专家何德全院士任编委会主任,以国家保密通信重点实验室为核心,集聚国内信息安全界知名专家学者,潜心数年编写的“网络与信息安全前沿技术丛书”即将分期出版。丛书有如下特点:一是全面系统。丛书涵盖了密码理论与技术、网络与信息安全基础技术、信息安全防御体系,以及近年来快速发展的大数据、云计算、移动互联网、物联网等方面的安全问题。二是适应面宽。丛书既很好地阐述了相关概念、技术原理等基础性知识,又较全面介绍了相关领域前沿技术的最新发展,特别是凝聚了作者

们多年来在该领域从事科技攻关的实践经验,可适应不同层次读者的需求。三是权威性好。编委会由我国网络和信息安全领域权威专家学者组成,各分册作者又均为我国相关领域的知名学者、学术带头人,理论水平高,并有长期科研攻关的丰富积累。

我认为该丛书是一套难得的系统研究网络信息安全技术及应用的综合性书籍,相信丛书的出版既能为公众了解信息安全知识、提升安全防护意识提供很好的选择,又能为从事网络信息安全人才培养的教师和从事相关领域技术攻关的科技工作者提供重要的参考。

作为特别关注网络信息安全技术发展的一名科技人员,我特别感谢何德全院士等专家学者为撰写本书付出的艰辛劳动和做出的重要贡献,愿意向读者推荐该套丛书,并作序。

何德全

协议是计算机网络和分布式系统中各种通信实体间相互交换信息时必须遵守的一组规则或约定,这些规则明确规定了所交换的数据格式及有关的同步问题,从而保证了通信双方有条不紊、可靠地交换信息。自从英格兰国家物理实验室(National Physical Laboratory, NPL)的 R. A. Scantlebury 和 K. A. Bartlett 在 1967 年第一次将“协议”(protocol)一词用于描述数据通信过程以来,已有大量通信协议出现并被标准化,广泛用于各种各样的网络和通信应用中。比较著名的网络协议有 TCP/IP 协议栈中的一系列协议,如 IP、TCP、UDP、POP3、SMTP、HTTP 协议等。

除了大量标准化的通信协议外,网络中还存在大量私有协议(也称为未知协议),各软件厂商或个人出于经济利益、安全、隐私等因素的考虑,并没有公开协议细节;一些恶意软件也采用了自己的私有协议防止被跟踪和分析。

网络协议规范对于网络管理、网络攻防有着重要意义。从网络管理的角度看,识别网络流量使用的传输协议对于提高网络服务质量、了解网络运行状态、监控恶意网络应用具有重要意义。从网络攻击的角度来看,假冒攻击、网络监听等主动和被动攻击技术都需要以协议格式为前提。从网络防护的角度看,识别软件使用的通信协议、分析软件间网络交互报文是软件安全性分析、漏洞挖掘、流量控制和网络安全策略制定等工作的重要内容。而对于迅速传播的蠕虫、僵尸程序、木马等恶意软件,快速分析其通信协议、掌握其命令控制方式更是对恶意软件做出及时反应的关键步骤。因此,以网络协议为主要研究对象的协议分析技术应运而生。

协议分析技术主要分为两大类,一类是对已知协议的识别与分析,另一类是对未知协议的逆向分析。前者以协议特征(如协议格式特征、端口特征、流量特征等)为基础,识别应用使用的通信协议并根据协议规范对协议报文进行分析,其前提是协议规范已知。后者则是在协议特征未知的条件下,通过协议报文或协议软件执行过程分析得到协议规范(包括协议

格式和协议状态机),即协议逆向分析。

协议逆向工程(Protocol Reverse Engineering),是指在不依赖于协议描述的情况下,通过对协议实体的网络输入/输出、系统行为和指令执行流程进行监控和分析,提取协议语法、语义和同步信息的过程,是工程化的协议逆向分析方法。人工方式的协议逆向虽然取得了较为理想的效果,但其过程冗长耗时、费力,且准确率依赖于分析人员的经验。随着网络规模的扩大和应用种类的增多,对协议逆向的准确性和时效性的要求越来越高,自动化的协议逆向分析技术成为人们追求的目标。

在江苏省自然科学基金项目“协议逆向工程关键技术研究”以及军队有关项目资助下,解放军理工大学协议逆向分析课题组经过多年研究,在自动化的协议逆向分析方面取得了一些研究成果。这些研究成果以及国内外在协议逆向分析方向的最新研究进展形成了本书的主体内容。

全书共8章,主要介绍协议逆向分析原理及应用技术。第1章绪论,主要介绍协议、协议逆向工程等相关概念、研究现状。第2章主要介绍协议设计原理,内容包括协议设计模型、内容、差错控制技术以及典型协议简介等,以此作为协议逆向分析的背景知识。第3章主要介绍协议规范描述模型,内容涉及协议逆向分析对协议规范描述模型的要求、著者提出的基于高阶属性文法的协议规范描述模型,这部分内容是一体化的协议逆向分析技术的基础。协议格式逆向分析主要有两种方法:基于网络流量的协议格式逆向分析和基于执行轨迹的协议格式逆向分析。第4章主要介绍基于网络流量的协议格式逆向分析,内容包括该方法的一般原理、研究现状、著者提出的基于递归聚类的协议格式提取方法、人工知识在逆向分析中的应用等。第5章主要介绍动态二进制程序分析技术,内容包括动态污点分析、动态符号执行技术以及常见的二进制程序分析平台,本章是基于执行轨迹的协议格式逆向分析技术的基础。第6章介绍基于执行轨迹的协议格式逆向分析方法,包括该方法的一般原理、研究现状以及作者在此方向上的研究成果。第7章介绍协议状态机推断技术,包括状态机推断原理、研究现状以及著者提出的两种状态机推断方法。第8章以针对网络协议的模糊测试技术为例,介绍网络协议逆向分析技术的应用。

本书是国内第一本全面介绍自动化网络协议逆向分析及应用的学术专著,可作为从事软件及协议逆向分析、网络管理、网络安全、网络对抗等方向的教学、科研及工程技术人员的参考书。

本书所展现的大量研究成果是著者吴礼发教授的博士研究生、硕士研究生得

到的,没有他们的创新性研究和勤奋努力,就不可能取得这些成果。此外,本书也引用了大量国内外网络协议逆向分析领域的学位论文和期刊论文中介绍的研究成果,在此不能一一列举他们的姓名,谨以此书的出版来表达著者对他们的感谢。

由于网络协议逆向分析涉及的内容比较广,加之著者水平的限制,书中难免存在缺点和错误,敬请读者批评指正。

著者

(wulifa@vip.163.com)

2015年9月于南京

目 录

第1章 绪论	1
1.1 协议	1
1.1.1 定义	1
1.1.2 网络体系结构	2
1.2 协议分析	8
1.2.1 应用需求	8
1.2.2 协议逆向工程	11
参考文献	14
第2章 协议设计原理	15
2.1 协议模型	15
2.2 协议设计的基本内容	17
2.2.1 协议的通信环境	17
2.2.2 协议提供的服务	18
2.2.3 协议功能	19
2.2.4 协议元素	22
2.3 差错控制技术	27
2.3.1 差错类型	27
2.3.2 差错检测技术	28
2.3.3 报文丢失、重复、失序处理技术	38
2.3.4 差错控制与层次的关系	43
2.4 典型协议	45
2.4.1 HDLC 协议	45
2.4.2 PPP 协议	49

2.4.3	IP 协议	52
2.4.4	TCP 协议	54
2.4.5	HTTP 协议	57
2.5	总结与展望	60
	参考文献	60
第3章	协议规范描述模型	61
3.1	概述	61
3.2	协议规范描述需求分析	62
3.3	高阶属性方法	64
3.3.1	协议格式分析	64
3.3.2	高阶属性文法	64
3.4	基于高阶属性方法的协议规范描述模型	67
3.5	模型实现	73
3.6	总结与展望	83
	参考文献	83
第4章	基于网络流量的协议格式逆向分析	85
4.1	概述	85
4.2	序列比对技术	86
4.2.1	双序列比对	87
4.2.2	多序列比对	88
4.2.3	问题分析	89
4.3	典型报文序列分析方法	90
4.3.1	PI	90
4.3.2	PEXT	91
4.3.3	BFS	94
4.3.4	ScriptGen	96
4.3.5	Discoverer	98
4.3.6	Automaton	100
4.3.7	Netzob	101
4.4	基于递归聚类的协议格式提取方法	103

4.4.1	基本块划分	104
4.4.2	递归分析	105
4.4.3	报文结构信息分析	110
4.4.4	语义及取值约束推断	110
4.4.5	算法复杂度分析	120
4.5	人工知识在逆向分析中的应用	121
4.5.1	人工知识	122
4.5.2	半自动协议逆向分析	122
4.6	RPRA 实现	127
4.6.1	输入模块	127
4.6.2	自动分析模块	130
4.6.3	输出模块	130
4.6.4	纠正模块	132
4.7	应用实例	135
4.7.1	样本集的获取	136
4.7.2	数据预处理	136
4.7.3	已知协议分析	136
4.8	总结与展望	148
	参考文献	149
第 5 章	动态二进制程序分析技术	151
5.1	概述	151
5.2	动态污点分析技术	152
5.2.1	动态污点分析的原理	152
5.2.2	污点属性的传播特征	154
5.2.3	动态污点分析技术的应用	155
5.3	动态符号执行技术	155
5.3.1	符号执行的原理	156
5.3.2	符号执行技术的局限	158
5.3.3	动态符号执行的原理及应用	159
5.4	二进制分析平台简介	161
5.4.1	Intel Pin	161

5.4.2	BitBlaze	169
5.4.3	其他二进制分析平台	178
5.5	总结与展望	181
	参考文献	182
第6章	基于执行轨迹的协议格式逆向分析	185
6.1	概述	185
6.2	典型指令序列分析方法	186
6.2.1	Polyglot	186
6.2.2	AutoFormat	188
6.2.3	Tupni	191
6.2.4	Prospex	193
6.2.5	ReFormat	194
6.2.6	Dispatcher	196
6.3	基于混合符号执行的协议格式提取方法	197
6.3.1	概述	197
6.3.2	基本思想	199
6.3.3	基于中间语言的混合符号执行技术	201
6.3.4	语义解析层次的协议格式提取技术	205
6.3.5	原型实现及应用	217
6.4	总结与展望	225
	参考文献	226
第7章	协议状态机推断技术	229
7.1	概述	229
7.2	基本定义	230
7.3	状态机推断技术研究现状	231
7.3.1	基于指令序列的状态机推断研究	231
7.3.2	基于报文序列的状态机推断研究	232
7.3.3	两类方法的比较	237
7.4	测试驱动状态融合的协议状态机推断方法	238
7.4.1	状态融合问题分析	238

7.4.2	方法概述	241
7.4.3	初始状态机构造	243
7.4.4	状态匹配与融合	245
7.4.5	状态融合验证	248
7.4.6	实例分析	249
7.5	基于域知识的协议状态机主动推断算法	256
7.5.1	概述	256
7.5.2	L_N^+ 算法	258
7.5.3	基于强顺序约束关系的 output query 过滤机制	259
7.5.4	基于 EPTT 的 output query 预响应机制	261
7.5.5	基于正例样本变异的 equivalence query 近似判定算法	262
7.5.6	实例分析	263
7.6	总结与展望	268
	参考文献	269
第 8 章	协议逆向分析的应用	273
8.1	概述	273
8.2	Fuzzing 测试技术	273
8.3	基于模型的 Fuzzing 技术面临的问题	279
8.3.1	数据格式的描述	279
8.3.2	测试用例的生成	281
8.4	文法驱动的 Fuzzing 测试技术	284
8.4.1	文法分析树的构造	286
8.4.2	测试节点的选择	290
8.4.3	基于语义的测试用例生成	291
8.5	应用实例	295
8.6	总结与展望	301
	参考文献	302
附录	缩略语	304