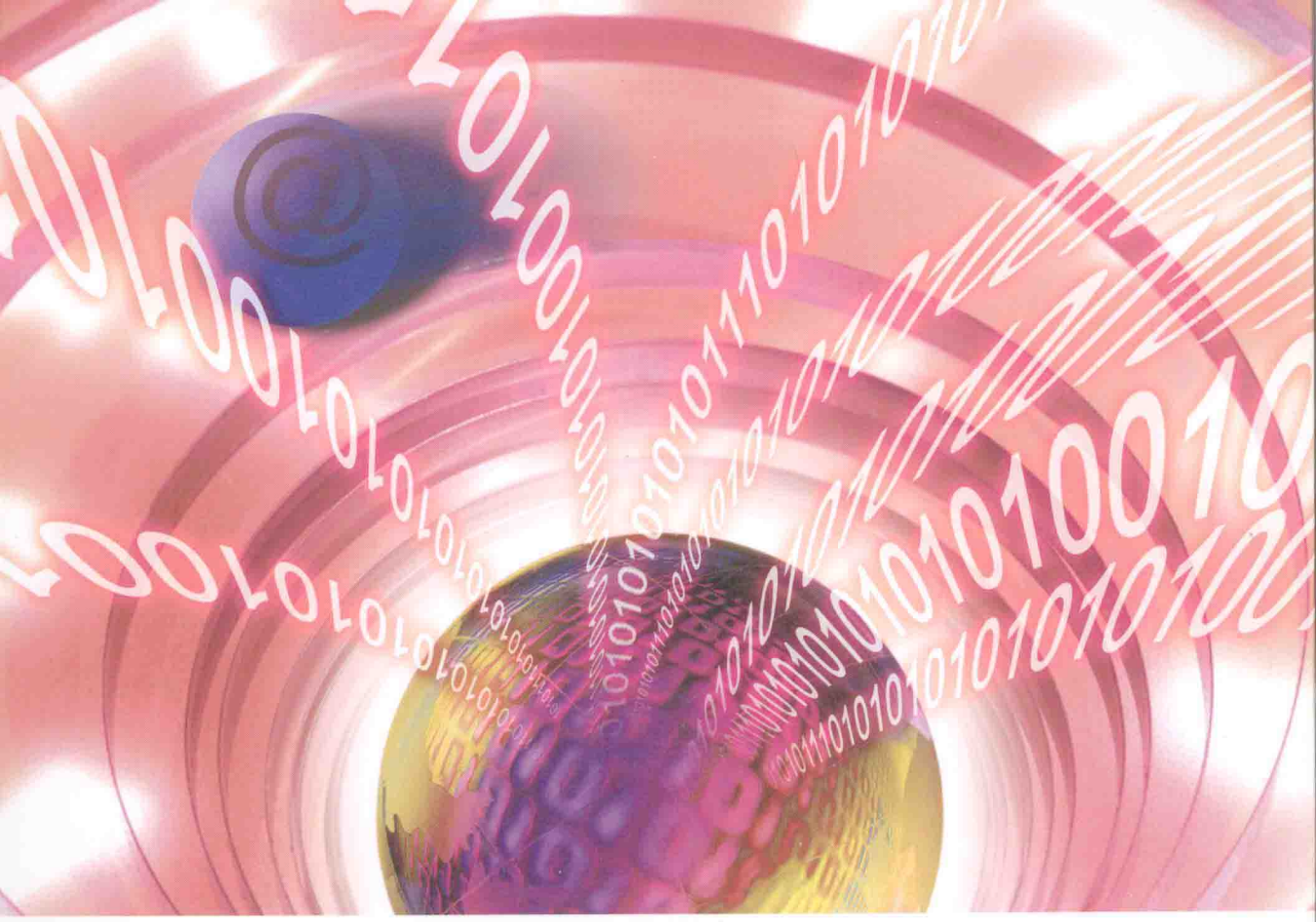


网络新技术系列丛书

JIANGSHI WANGLUO
JIANCE JISHU

僵尸网络检测技术

程光　吴桦　王会羽　张军　陈玉祥◎著

东南大学出版社
SOUTHEAST UNIVERSITY PRESS

网络新技术系列丛书

僵尸网络检测技术

程 光 吴 桦 王会羽
张 军 陈玉祥 著

东南大学出版社
SOUTHEAST UNIVERSITY PRESS
·南京·

内 容 提 要

僵尸网络是指通过各种手段在多台计算机中植入恶意程序，使僵尸控制者能相对方便和集中地控制这些计算机，向这些受控制的计算机发布各种指令进行相应恶意活动的攻击网络。本书首先通过大量实验介绍在僵尸和人的活动行为方面的差异，其次介绍一种基于主动技术的恶意代码捕获方法以及获得僵尸样本的通信数据和行为特征的方法，最后介绍6种不同类型的僵尸网络检测方法。全书包括10章内容，第1章综述了近年来国内外主要的僵尸网络检测方法的研究进展；第2章介绍僵尸的会话行为分析；第3章介绍主动僵尸代码捕获方法；第4章介绍僵尸提取和分析；第5章介绍僵尸源码实例分析；第6章介绍基于DNS的僵尸网络检测方法；第7章介绍基于C&C信道的僵尸网络检测方法；第8章介绍基于流量行为的僵尸网络检测方法；第9章介绍基于事件关联的僵尸检测方法；第10章介绍基于内容解析的僵尸网络检测方法。

本书可供计算机科学、信息科学、网络工程及流量工程等学科的科研工作人员、大学教师和相关专业的研究生与本科生，以及从事计算机网络安全领域、网络工程及网络管理的工程技术人员阅读参考使用。

图书在版编目(CIP)数据

僵尸网络检测技术/程光等著. —南京：东南大学出版社，2014.10

网络新技术系列丛书

ISBN 978-7-5641-4945-1

Ⅰ.①僵… Ⅱ.①程… Ⅲ.①计算机网络—安全技术 Ⅳ.①TP393.08

中国版本图书馆CIP数据核字(2014)第099523号

僵尸网络检测技术

出版发行	东南大学出版社
出 版 人	江建中
社　　址	南京市四牌楼2号
邮　　编	210096
经　　销	全国各地新华书店
印　　刷	江苏凤凰扬州鑫华印刷有限公司
开　　本	787 mm×1092 mm　1/16
印　　张	15.75；彩插8面
字　　数	403千字
书　　号	ISBN 978-7-5641-4945-1
版　　次	2014年10月第1版
印　　次	2014年10月第1次印刷
定　　价	40.00元

（本社图书若有印装质量问题，请直接与营销部联系，电话：025—83791830）

引 言

互联网的出现极大地改变了人们的生活方式和工作习惯，给人们带来了各个方面的便利，已经成为人们共享信息和交流的重要平台和工具。但是同时互联网中网络安全事件的发生也越来越频繁，如分布式拒绝服务攻击、垃圾邮件发送、网络钓鱼、蠕虫传播、敏感信息窃取等，这些恶意行为给社会造成了巨大的经济损失，而且使全球互联网的网络安全领域面临着前所未有的挑战。

僵尸网络是指通过各种手段在多台计算机中植入恶意程序，使僵尸控制者能相对方便和集中地控制这些计算机，向这些受控制的计算机发布各种指令进行相应恶意活动的攻击网络。通过僵尸网络，攻击者(Attacker)可以迅速地控制数以万计的主机，导致僵尸网络的规模呈指数级增长。僵尸网络现在已经成为了网络攻击者的首选平台，是操纵这些恶意行为的“幕后指使者”。如此大规模的僵尸网络被掌握在攻击者手中，不仅会对整个互联网的安全带来隐患而且会对经济社会产生严重的损害，甚至会产生违法犯罪活动。当前所有的 DDoS 攻击都是由僵尸网络平台发起的，95%以上的垃圾邮件也是由僵尸网络发送的，其中 2013 年 9 月的垃圾邮件数量接近了 4 万亿封，同时僵尸网络特别擅长信息窃取和钓鱼攻击。由此可见，僵尸网络已经成为危害互联网的重大安全威胁之一，严重地影响了人们的生活、社会的安定和经济的发展，如何有效地检测、追踪和控制僵尸网络已经成为安全领域的重点。

命令与控制机制是区分僵尸网络的主要依据，也是保证僵尸网络正常工作的前提条件和有效途径。只有理解了命令与控制的实现机制，才能深入理解僵尸网络的工作机理。按照控制信道使用互联网协议的不同，僵尸网络可分为 IRC 僵尸网络、HTTP 僵尸网络、P2P 僵尸网络。僵尸网络现在逐渐地融合了其他各种恶意代码的技术，包括 Rootkit 隐藏技术、加密技术、蠕虫传播技术、网站挂马技术、混淆技术等，这使得新出现的僵尸网络具有更加强大的破坏力和隐蔽性。

在僵尸和人的活动行为差异方面，本书首先介绍对一个互联网聊天服务商的 21 个不同的聊天室搜集 2 个月聊天日志，从日志中识别 14 种不同的聊天 bot 且将它们分为四类，通过对聊天 bot 和人类用户的消息间隔与消息尺寸的统计分析，发现 bot 与人类用户的行为的区别，同时介绍了使用主动探测技术检测僵尸网络 C&C 通信的方案。

在僵尸样本的获取与原理分析方面，本书首先介绍一种基于主动技术的恶意代码捕获方法，通过结合网络爬虫技术和高交互式客户端蜜罐技术，实现了自动化恶意代码的捕获。其次，利用反病毒引擎扫描恶意软件，获取僵尸样本，再通过虚拟机自动化运行僵尸样本，得到僵尸样本的通信数据和行为特征。并且针对 SDBot 和 ZeuS 僵尸程序做源码分析，这些分析为深入研究 IRC 和 HTTP 僵尸网络奠

定了基础。

在僵尸网络检测技术方面，本书介绍了5类检测技术。根据僵尸请求DNS流量的周期性、群体性等特性，介绍了僵尸网络的贝叶斯检测方法；根据僵尸网络C&C信道的时空相关性和相似性特性，介绍了一种异常检测方法和一种基于聚类的无监督检测方法；通过研究SDBot、ShadowBot、BlackEnergy、ZBot 4种僵尸客户端与C&C服务器交互时产生的流量特性，以流序列为研究对象，针对性地提取流间隔时间、流持续时间、流字节数、交互频率等特征，采用一维聚类及相应的评分方式，建立僵尸模型并与未知流量进行匹配；由于僵尸网络客户端往往会根据僵尸控制者的指令在特定的时间执行特定的恶意行为，因此介绍了两种基于行为的僵尸网络检测方法和一种关联分析算法；根据僵尸流量中的具体特征内容，介绍了一种通过识别和解析IRC协议以获取IRC协议的信道和昵称信息的检测方法，以及基于已知的C&C通信实例和部署了模板的网络的流量的检测方法。

全书包括10章内容，第1章综述了近年来国内外主要的僵尸网络检测方法的研究进展；第2章介绍僵尸的会话行为分析；第3章介绍主动僵尸代码捕获方法；第4章介绍僵尸提取和分析；第5章介绍僵尸源码实例分析；第6章介绍基于DNS的僵尸网络检测方法；第7章介绍基于C&C信道的僵尸网络检测方法；第8章介绍基于流量行为的僵尸网络检测方法；第9章介绍基于事件关联的僵尸检测方法；第10章介绍基于内容解析的僵尸网络检测方法。

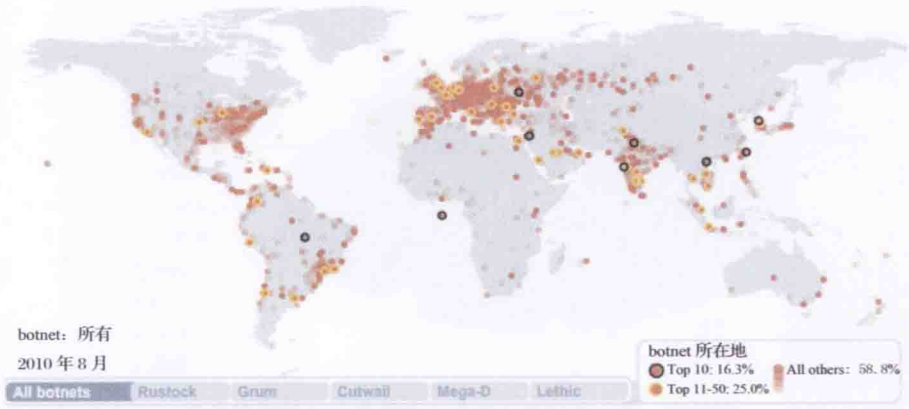
本书主要是作者在网络安全和僵尸网络检测分析领域长期的研究成果的总结，也包括了翻译总结的部分国际上僵尸网络检测方面最经典的科技研究论文。全书由程光统稿完成，在本书撰写过程中，张阳、郭晓军、王玉祥、周爱平、程志、相增辉、潘吴斌、梁一鑫、陈玉祥、汪志、蔡振盛等也对本书的工作给予了大力支持，参与了本书部分章节的编写工作以及本书的整编和校验工作，在此表示感谢。感谢东南大学龚俭教授、丁伟教授、杨望讲师、吴珺助工以及解放军理工大学的张国敏等人的帮助和支持。

本书研究成果受江苏省科技支撑计划——工业部分：基于分布式通信机理的主干网僵尸网络追踪系统(No. BE2011173)；国家973计划：基于自治治理模型的网络管理与安全研究(No. 2009CB320505)；国家自然科学基金：高速网络活跃节点检测及其流量分类研究(No. 60973123)；江苏高校优势学科建设工程资助项目(PAPD)等国家省部级项目的资助，在此表示感谢！同时感谢江苏省“六大人才高峰计划”(No. 2011—DZ024)的支持。

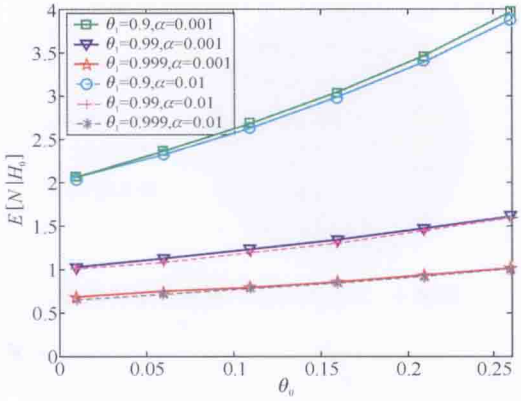
在本书的撰写过程中，得到了东南大学计算机科学与工程学院、计算机网络和信息集成教育部重点实验室(东南大学)、东南大学出版社等单位领导和专家的大力支持，在此深表谢意！同时对作者所引用的参考文献的作者及不甚疏漏的引文作者也一并致谢！由于作者水平有限，编写过程中难免存在很多不足及顾此失彼之处，敬请读者给予批评指正！

著　者

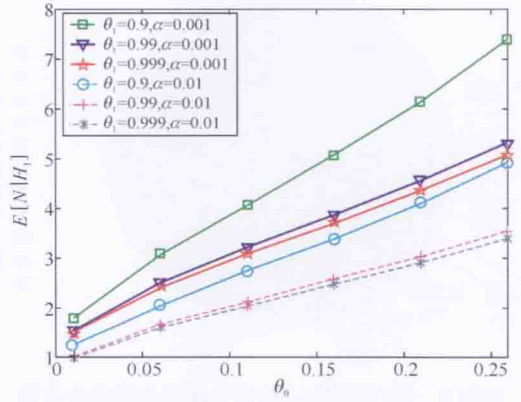
2014年2月



彩插 1　僵尸网络全球分布情况

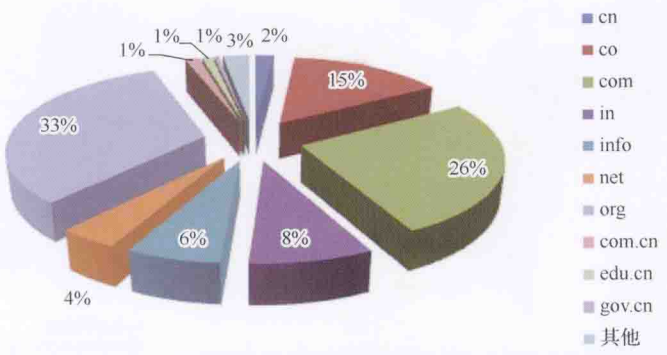


（a）检测人类用户 IRC 平均轮数

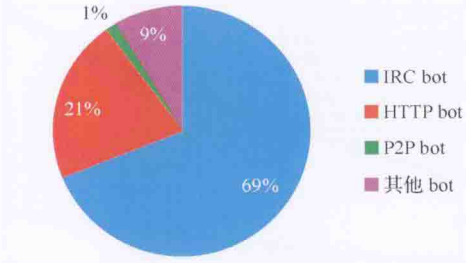


（b）检测僵尸用户 C&C 平均轮数

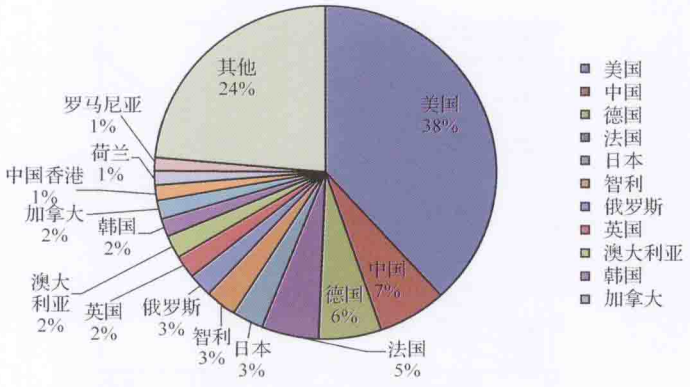
彩插 2　对人类用户的感染以及对检测的影响



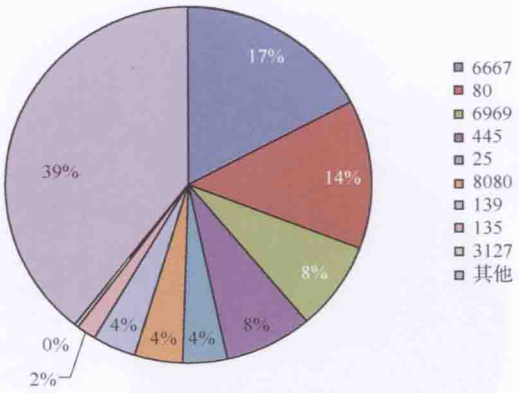
彩插 3　2011 年中国大陆地区恶意网站按域名分布图



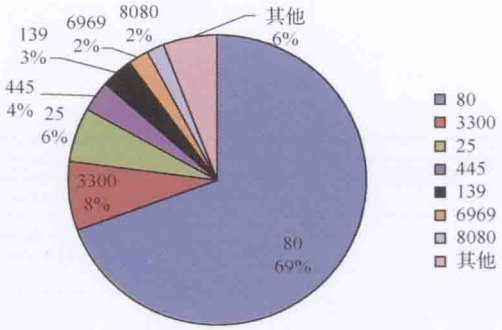
彩插 4　三种类型僵尸程序数量比例图



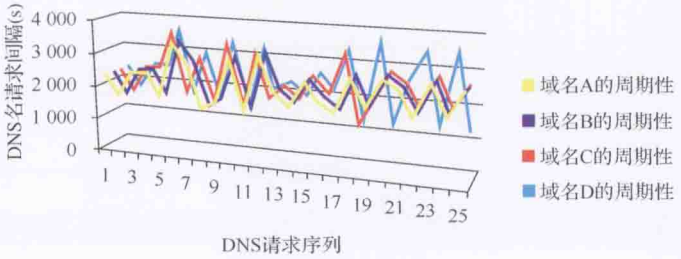
彩插 5　僵尸网络 C&C 服务器所在国家分布图



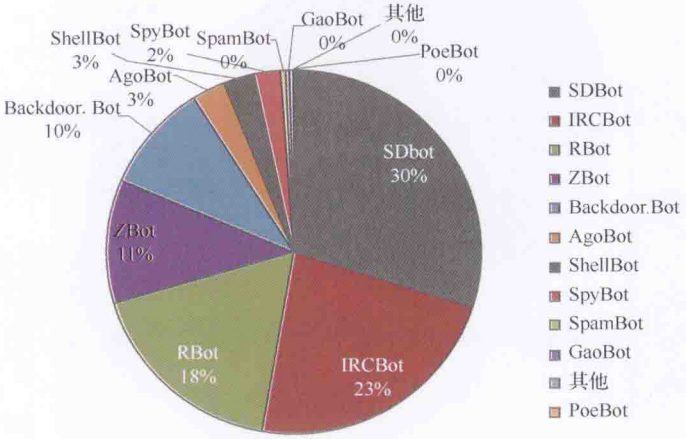
彩插 6　IRC 僵尸网络控制服务器端口分布图



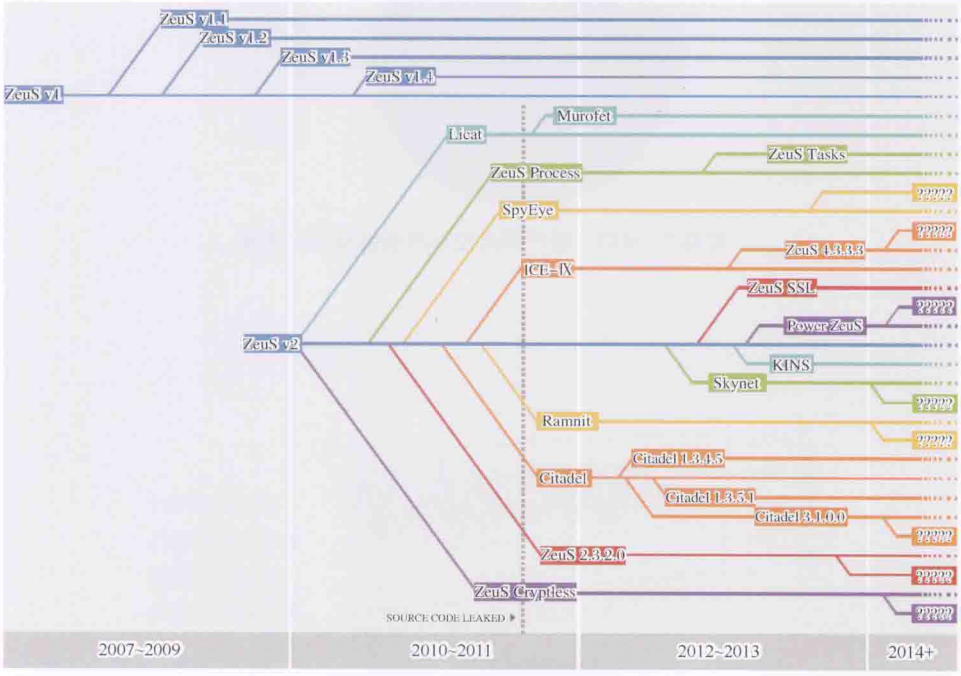
彩插 7　HTTP 僵尸网络控制服务器端口分布图



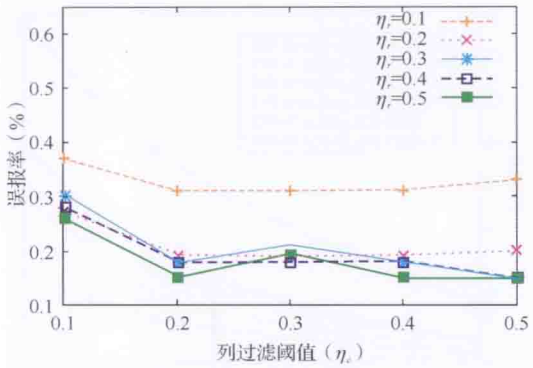
彩插 8　Backdoor. Bot. 160896 对 4 个域名的 DNS 请求的周期性比较



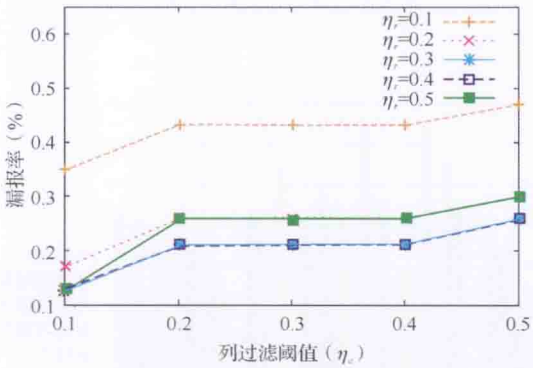
彩插 9　黑名单中的 IP 地址对应的僵尸类型



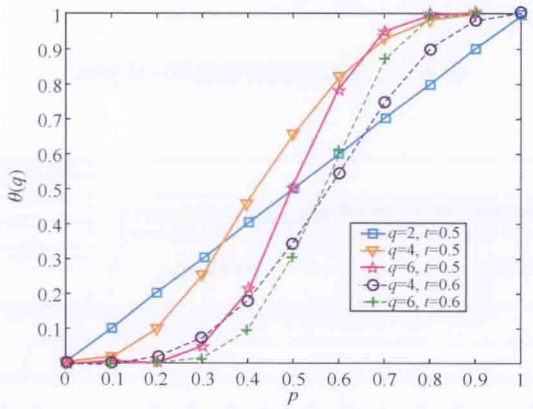
彩插 10　ZeuS 版本演变历程



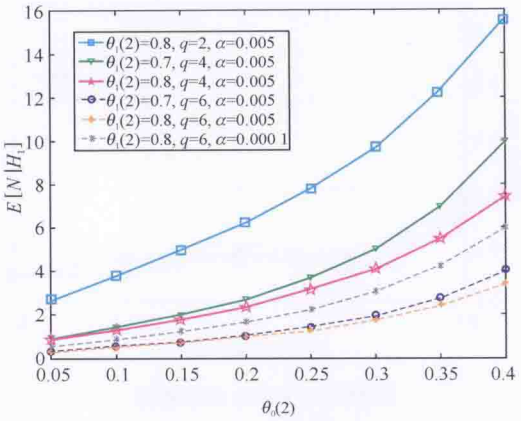
彩插 11　trace1 的误报率



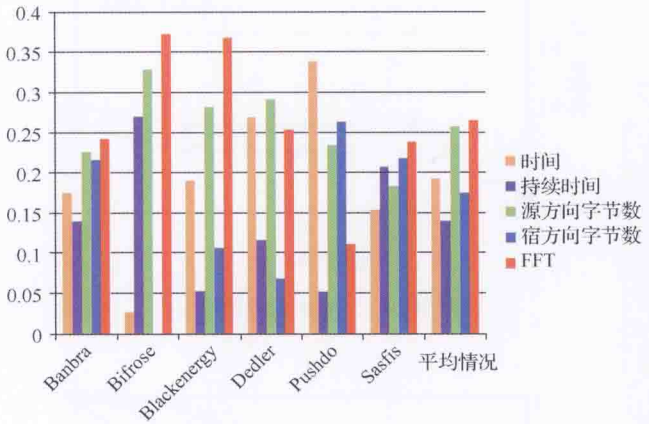
彩插 12　trace1 的漏报率



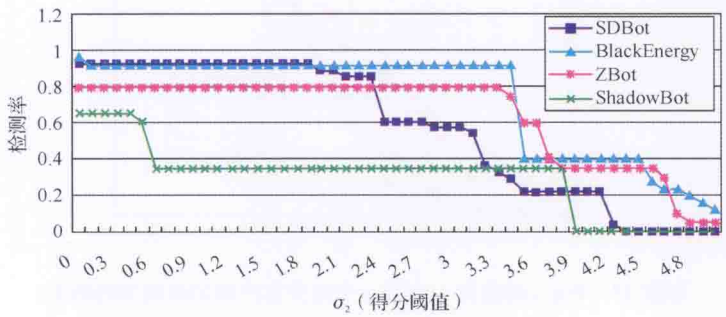
彩插 13　$\theta(q)$，阈值为 t，含有 q 个相关客户端的群均匀的概率



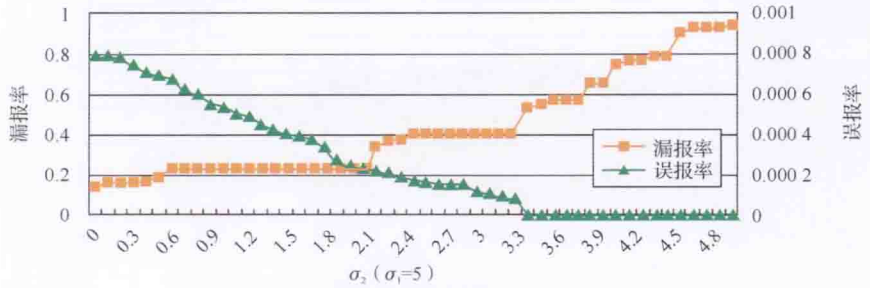
彩插 14　$E[N|H_1]$，僵尸网络情况下响应轮数的期望（改变 $\theta_0(2)$，q，α，设 $\beta=0.01$）



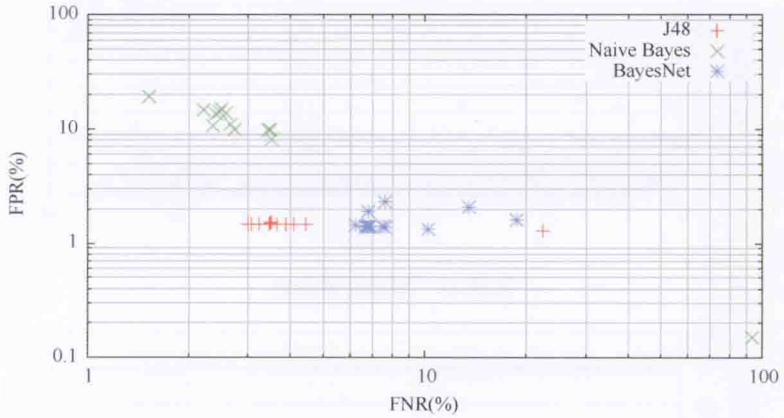
彩插 15　成功识别特征的归一化贡献



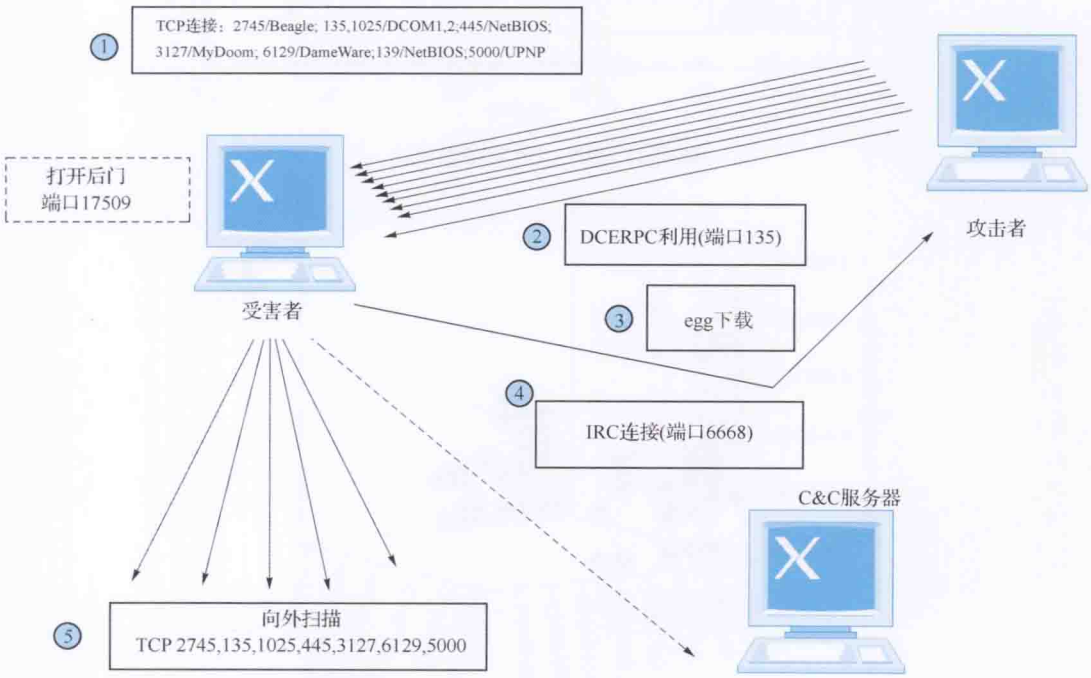
彩插 16　4 类 bot 检测效果比较（$\sigma_1=5$）



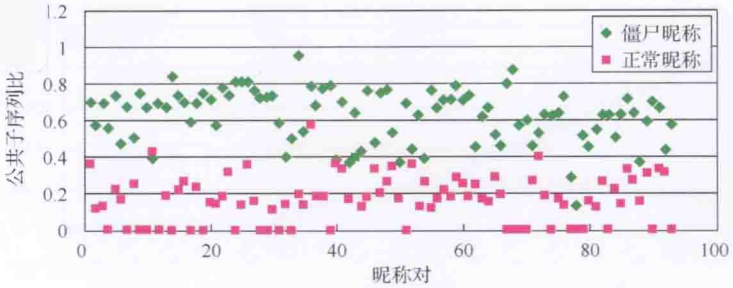
彩插 17　系统整体检测效果



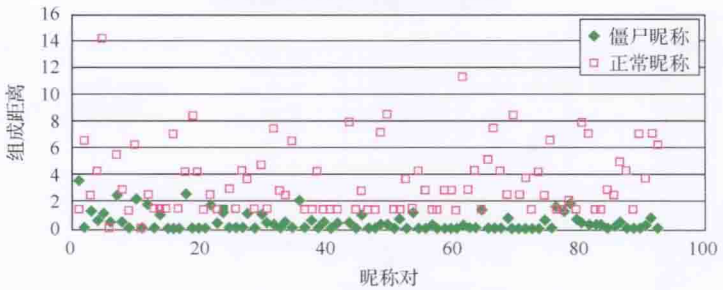
彩插 18　三种方法对标记的 X 建筑 trace 的 FNR 和 FPR 实验结果



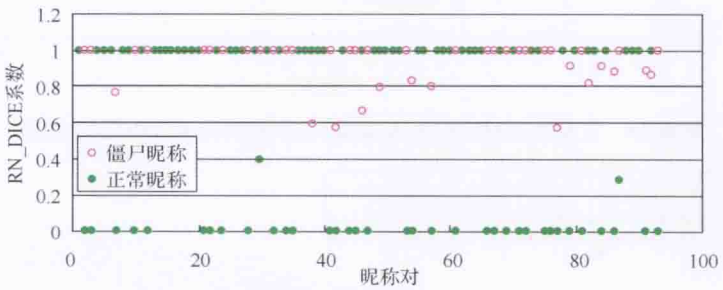
彩插 19　PhatBot 会话描述



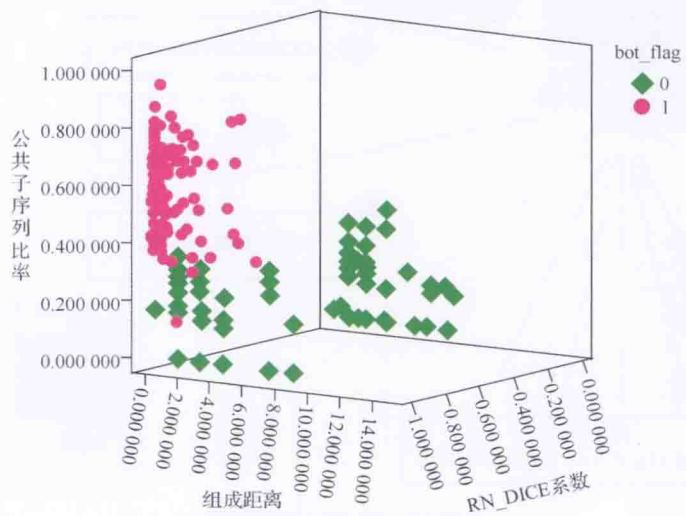
彩插 20　公共子序列比率对比



彩插 21　组成距离对比



彩插 22　RN _ DICE 系数对比



彩插 23　僵尸昵称/正常昵称分布对比

目　录