doi:10.3969/j.issn.1000-436x.2014.z1.011

恶意代码自动分析系统的研究

赵毅，龚俭，杨望
（东南大学 计算机科学与工程学院，江苏 南京 211189）

摘 要：恶意代码的网络行为分析是网络安全领域的一个重要研究视角。针对现有系统普遍存在的网络行为分析不全面、不深入的问题，归纳了恶意代码的功能模块，提出了较为全面的网络行为分析内容。通过对比已有系统的网络行为分析功能，选取合适的系统 CUCKOO 作为基础平台。通过实例对其网络行为分析功能进行详细分析，并提出了优化、扩展方案。

关键词：网络安全；恶意代码；自动分析
中图分类号：TP309 **文献标识码**：A **文章编号**：1000-436X(2014)Z1-0052-06

Study on modern malware analysis system

ZHAO Yi, GONG Jian, YANG Wang
(School of Computer Science and Engineering, Southeast University, Nanjing 211189, China)

Abstract: The analysis of malicious code's network behavior is an important research field of network security. This function of existed systems is incomplete and not deep. The functions of malicious code are summarized and a comprehensive content is presented. Moreover the network behavior analysis function of existed analysis systems is introduced and CUCKOO which is able to satisfy the requirements of involved study is found. Finally the advantage and points of this application platform were summarized, and an expansion of the system was proposed.

Key words: network security; malware; automated analysis

1 引言

随着网络基础资源、用户数量的迅猛增长，以及网络技术、网络应用的蓬勃发展，网络资源逐渐成为社会赖以运转的核心基础设施，已经成为目前各个组织争夺的新战场，高级网络渗透技术成为重点发展的对象。恶意代码作为主要的技术载体和实施手段，自然成为了网络安全领域的热点。

恶意代码的网络行为分析是该领域一个重要研究视角，主要包括恶意代码通信协议的协议格式逆向分析、协议模型推理网络流捕获与分析和基于网络流的检测特征自动提取方法等。

目前，文献不能给出恶意代码严格的统一定义，研究者往往根据问题研究的需要从不同的角度进行刻画，或者依据某恶意代码样本的主要功能为其打上类别标签，用于描述某种典型的恶意行为，例如僵尸代码。本文的重点即为僵尸代码的网络行为分析。

收稿日期：2014-10-14

2 僵尸代码的网络行为及其功能

本节介绍僵尸代码网络行为的功能，论述关注恶意僵尸网络行为分析的重要意义。

整体看来，需要通过网络通信实现其功能的恶意代码占有很高比例。Bayer 等[1]在基于动态分析环境的大规模恶意代码研究中发现：55.18%的样本在执行时表现了网络活动，部分具有网络通信能力的样本由于执行条件限制没有表现网络活动[2]。不同类型的恶意代码通过网络通信完成的功能各有不同。

僵尸代码的一大特征便是必须要与其 C&C 进行通信。通信的主要目的是获得攻击者的命令，根据命令要求完成相应功能，如获得受害主机信息、下载配置文件、扫描局域网主机信息、发起 DDoS 攻击、下载垃圾邮件模板等，并根据需要更新其功能组件。诸葛建伟等[3]在僵尸网络的发

现与跟踪中对僵尸功能进行了分类，本文从通信行为的视角进行调整并对可能产生通信的模块进行了标注，如表 1 所示。

表 1　　僵尸程序模块分类

功能模块	模块分类		
主体功能模块	僵尸主机管理模块	注册	●
		管理	●
		删除	●
	僵尸主机控制模块	DDoS 攻击	●
		架设服务	
		PC 控制	
		发送垃圾邮件	●
		代理	●
		命令执行	
辅助功能模块	网络传播模块	漏洞攻击	●
		现有后门	●
		蠕虫木马	●
	下载与更新模块	下载	●
		更新	●
		访问	●
	躲避检测与分析模块	变形多态	
		代码混淆	
		反虚拟	
		防杀毒	
		反调试	
	息窃取模块	系统信息	●
		网络环境	●
		邮件列表	●
		密码	●
		软件密钥	●

表 1 中的“●”为需要进行网络通信的功能模块。可以看出，除了躲避检测与分析模块，僵尸代码的绝大多数功能模块的实现依赖于网络通信。

3　网络行为分析的应有内容

本节依据僵尸网络相关领域的研究文献，总结文献中需要的研究素材作为需求来源，提出较为全面的僵尸代码分析内容,并结合实例进行介绍。

3.1　僵尸代码网络行为分析的主要内容

对僵尸网络的研究可归纳为检测、追踪、测量、预测和对抗 5 个部分。根据命令与控制信道使用的控制协议不同其检测、追踪、测量、预测和对抗方法均有不同。针对 IRC 协议的相关研究[4~8]，其素材主要包括 IRC 信道中的用户昵称、用户数量、IRC 命令的数量和频率、句子中平均单词数以及连接服务器失败的重连时间间隔等。针对 HTTP 协议的研究文献[9,10]，其素材主要包括周期性重联的度、IP 地址、端口等。针对 DNS 协议的研究文献[11~14]，主要涉及 DNS 请求的域名、请求速率、IP 地址、成功率等。

本文总结相关文献研究中涉及的素材，在表 2 中列出了比较全面的僵尸代码网络行为分析的内容，期望网络行为分析的结果更具实际意义。其中某些项如报文长度等针对单个样本时并没有太大价值，但当分析样本集达到一定规模便可以满足相关的统计研究。

表 2　　恶意代码网络分析的内容

协议	内容
IP	IP 地址
	URL
ICMP	TYPE
	CODE
IGMP	TYPE
TCP	源端口
	目的端口
	报文长度(length)
UDP	源端口
	目的端口
	报文长度(length)
DNS	域名(name)
	请求类型(query type)
	请求结果(query result)
	是否成功
	传输层协议(protocol)
IRC	昵称
	频道
	IRC 命令
HTTP	url
	目的 IP+被叫端口
	状态（status）
	实体类型（content-type）
	shellcode
	additonal-malware
	超链接 url
	控制命令
SMTP	邮件账户
	邮件模板

3.2 典型僵尸代码的网络行为分析实例

Zeus 是目前非常知名的僵尸网络，用来窃取银行凭证及其他个人信息。2007 年首次被发现，2011 年 Zeus 源码泄露，致使其变种爆炸式增长，更有黑客为其添加 P2P 模块，对互联网的危害很大，影响深远。

本文以 Sourcefire VRT Labs[15]给出的 sourcefire 公司的主机感染 zeus 之后的通信数据为实例，对僵尸代码的网络行为分析进行介绍。

设备均使用 WinXP，service pack2，没有打补丁，通信集中于僵尸主机管理模块和下载与更新模块。在毫秒级的感染过程中，主机会发出 HTTP-GET 请求，导致 C&C 会发送一堆标记 Content-Type 为 application 或者 octet-stream 的二进制数据。首先，看似模糊的二进制数据，但 Zeus Tracker 网站的一些分析指出这是加密后的配置文件；其次，数据的大小变化巨大，例如，在给出的 3 段 pcap 中数据分别是 379 335 164 和 185 133 字节；再次，设计的 URL 变化很随机，限于篇幅，表 3 中列出了一部分。

表 3 Zeus 涉及 URL 示例

/Gallery/IMAG0081,GIF	/kartos/kartos,bin
/btn001/config,bin	/ribbn,tar
/bugzy/i,cfg	/test/config,bin
/cnf/trl,jpg	/sell,jpg
/dzen/misc,inc,php	/ukk/cfg,bin
/film/video,bin	/z_bot/what,bin
/ftr/vosmoipont,bin	/zeus/config,bin
/ii1IGh,aeL8uf	/inmake/lds/cfg,bin
/index_files/4jpg,bin	/zs/cfg,bin

Zeus 除此之外的数据交换不同：一部分样本没有产生通信，一方面是逃避检测，另一方面 C&C 服务器暂时没有任务，也没有需要上报的用户敏感数据，或者 C&C 已经被政府部门、运营商关停；还有很大比例的样本下载了 Windows 下的 PE 文件，对客户端进行更新，PE 文件的名称也有很大的随机性，情况如表 4 所示。

表 4 Zeus 下载的 PE 文件名

/40,exe	/load,exe
/fshit/d,exe	/money-s2,exe
/good/tlz/server32,exe	/money,exe
/gus/windir,exe	/rhueirh4furh74,exe
/inmake/lds/server32,exe	/ser,exe
/kartos/krt,exe	/t,exe
/ldr/ldr,exe	/trhr7y4urjhe83,exe

4 使用现有分析系统的网络行为分析

由于恶意代码编写者与分析者之间对抗活动的持续演进，一方面，现代恶意代码在结构和属性上出现了新的特性，具备精巧的逃避分析策略、复杂的代码结构和多样的网络活动，例如，针对静态分析的反汇编技术，如内核、多态、变形、代码混淆（加壳、加密、混淆）、反虚拟、反调试等技术。

恶意代码分析系统也针锋相对地提出了许多可行的解决方案。从实现方法所处的位置来看，动态恶意代码分析技术的实现策略包括调试器环境、沙箱环境、基于虚拟机的分析平台和基于全系统模拟器的分析平台。

下面就结合上述实例和第 3 节给出的分析需求来探讨网络分析功能的扩展及优化。该分析文件的 MD5 值为 0c5e9f564115bfcbee66377a829de55f，其网络活动的基本情况为：首先对 johnford985. app-spot.com 进行 DNS，获取访问的 IP 地址 74,125, 137,141，然后与该地址的 443 端口进行了通信，详细情况如表 5 所示。

表 5 分析文件网络通信情况

源端口	发送字节	接收字节
1028	840	2 127
1029	697	446
1030	697	446
1031	697	150

表 6 列出了目前流行的 3 种具有代表性的不同层次的开源系统，分别对其网络行为进行分析。

表 6 开源分系统情况

系统名称	研发单位	分析平台	是否开源	版本	更新情况
TWMAN	台湾高速网路与计算中心	物理主机	是	版本 2,0	2013,6
CUCKOO	谷歌编程夏令营	Vbox/vmware	是	版本 1,1	2014,4
BitBlaze	加州大学伯克利分校	QEMU	是	版本 1,0	2010,6

TWMAN[16]是源自 Truman 的开放原码的自由软件，利用实体设备来构建分析环境，使用 client-server 结构来完成对整个分析流程的控制，基本流程如图 1 所示。

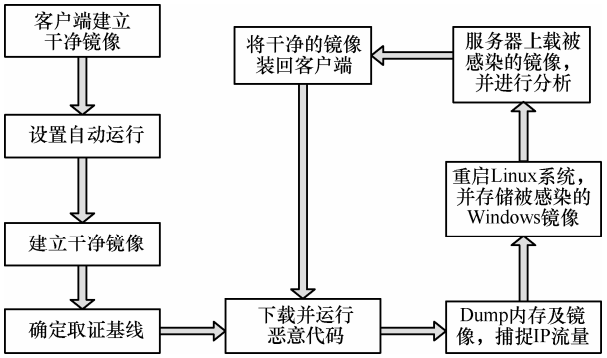

图 1　TWMAN 基本流程

TWMAN 的新版本中增加了网络流量的统计功能，分别对协议、目的 IP、端口进行统计，但也仅仅停留在计数，并未体现报文长度、时间间隔等更深层次的分析；增加了 DNS 查询但仅记录域名，并未体现 DNS 的其他属性。TWMAN 可以有效地反制新型态恶意程序所发展的 Anti-VM 技术，较其他系统硬件设备要求较高，平台部署步骤也稍显繁琐。但利用实体设备来构建分析环境，使分析环境的扩展受到硬件环境的诸多限制，并且一次只能进行一个恶意代码样本的分析，分析效率有很大的提升空间。

BitBlaze[17]的动态分析部件 TEMU 是基于全系统模拟器 QEMU 实现， QEMU 是一款使用动态可移植翻译技术的机器模拟器[18,19]，能在不需要做任何修改的目标操作系统上运行并且它集成一个 Linux 特有的用户模式模拟。这种运行模式可以用来测试交叉编译的结果或者可以用于测试 CPU 模拟器而不需要去启动一个完全的虚拟机。TEMU 是对 QEMU 的改造和扩展，以插件的方式来实现用户自定义分析工具的功能 ，其提供一系列接口，用户使用这些接口编写特定的 TEMU 插件，并在运行时加载插件，控制平台执行特定的分析任务。由于使用全系统模拟实现的动态执行环境，在执行效率上不可避免地与真实物理平台存在较大差异，特别是在某些分析要求下实施细粒度分析时，每条指令的执行都会引入额外的分析代码，使程序的执行效率大大降低。

CUCKOO 系统是基于虚拟机的分析平台，以虚拟机 KVM、VMware 和 Virtualbox 环境为依托，基于虚拟机监视器（VMI, virtual machine inspection）实现对虚拟机的 Guest 系统中运行的程序进行监控和检查，CUCKOO 基本流程如图 2 所示。

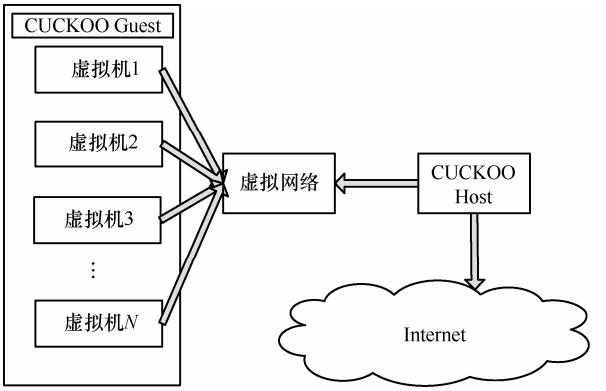

图 2　CUCKOO 基本流程

CUCKOO 系统的网络分析报告中仅给出了相关的域名、主机 IP、DNS 请求（仅给出请求域名与对应 IP）,但没有针对协议、IP、端口的统计功能。由于会给出捕获的 PCAP 文件，只需稍加改进便可实现统计功能。CUCKOO 从其使用 Virtualbox 作为分析支持的情况来看，虽然在反虚拟技术有弱势，可能会出现网络行为捕获不全的情况，但可以同时开启 5~10 台虚拟机作为分析环境，大大提高了其分析效率。

已有文献[20~23]给出的恶意代码分析系统具有良好的综合分析能力，但这些工具同样对网络行为分析没有给予足够关注。

从以上分析情况看，目前的系统在网络行为分析方面与第 3 节提出的分析需求相差甚远，需要选择合适平台进一步扩展、优化其网络行为的分析功能。

虽然 CUCKOO 基于虚拟机的分析平台在记录程序的指令轨迹方面有诸多不便，反虚拟能力弱于 TWMAN，但有很好的分析效率，也有一定的分析检测的能力，尤其是在网络行为分析方面有着很好的分析能力与效率。着眼于分析恶意代码的网络通信行为，以及对分析效率的要求，本文选定 CUCKOO 系统作为基础平台。

5　CUCKOO 分析系统的其他功能

在第 4 节通过对现有分析系统的对比分析，选取 CUCKOO 系统作为基础平台，本节结合实验详细介绍 CUCKOO 平台的基本功能，并给出了 CUCKOO 系统在关注恶意代码网络行为分析的视角下的可优化和扩展项。

CUCKOO 可以作为单独系统应用，同时由于其模块化的设计可以集成在更大的框架中，可以用于分析文件类型，如表 7 所示。

表 7　CUCKOO 系统可分析文件类型

Windows 下可执行文件
DLL 文件
PDF 文件
OFFICE 文件
URL 和 HTML 文件
PHP 脚本文件
CPL 文件
VB 脚本文件
ZIP 文件
JAR 文件
其他文件

本文安装、配置、调试 CUCKOO 系统正常后，提交从 Virusshare 获取样本，提交系统进行分析，限于篇幅，选取典型实例进行简要介绍，其分析结果目录如图 3 所示。

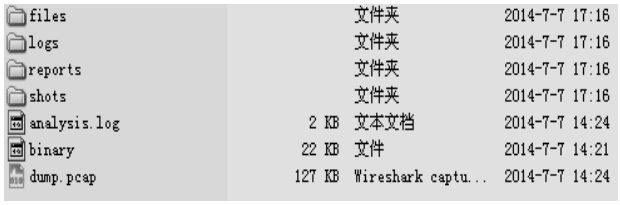

图 3　Cuckoo 系统分析结果目录

如图 3 所示，其中 files 文件为在系统中产生的文件，本实例中为 autoexec,bat 文件；logs 文件夹存放分析日志文件；shots 文件夹存放客户分析机的截屏图片；reports 用于存放产生的分析报告，包含 HTML、JSON 2 种格式；binary 为拷贝的分析文件副本；dump、pcap 文件为捕获的网络数据，其中包含 CUCKOO 主机与客户机的流量。在分析报告中包含的信息如表 8 所示。

对比第 3 节给出的恶意代码网络行为分析的内容可在以下方面对 CUCKOO 的网络行为功能进行扩展。

第一，可以对其分析的广度进行扩展，增加其网络行为分析项，如表 9 所示。

表 9　CUCKOO 系统可扩展的分析项

协议	内容
ICMP	TYPE
	CODE
TCP	报文长度（length）
Udp	请求类型(query type)
	请求结果(query result)
	是否成功
	传输层协议(protocol)
IRC	昵称
	频道
	IRC 命令
HTTP	shellcode
	additonal-malware
	超链接 url
	控制命令

第二，可以对其网络分析深度进行扩展。例如 DNS 相信分析绝大多建立、健全域名与 IP 之间的联系。绝大多数的恶意软件使用域名解析系统（DNS）进行地址解析。域名系统能够提供弹性和可持续性的服务，大多数合法商业组织并不经常更新自己的域名注册信息或者 DNS 记录。在实验过程发现一些域名表现出 fast flux 特征。例如在实验数据中，虽然只间隔数 10 min，但其 DNS 的返回结果并不相同，如表 10 所示。

表 8　分析报告中的基本内容

基本信息	签名信息	截屏信息	静态分析	文件信息	网络分析	系统行为
文件名、大小、MD5、PEiD、Yara、VirusTotal	本文件未匹配	系统截屏	Sections、Imports、Strings	文件名、大小及其他详细信息	相关主机信息、DNS 信息	Files、Registry Keys、Mutexes

表 10　恶意代码的 DNS 分析结果

VirusShare_0b680e7bd5c0501d5dd73164122a7faf	www.rbaparts.com	205.178.189.129
	www.rbaparts.com	205.178.189.131
VirusShare_0c5e9f564115bfcbee66377a829de55f	Johnford985.com	74.125.137.141
	Johnford985.com	74.125.136.141
VirusShare_0ca6e2ad69826c8e3287fc8576112814	www.fbrshop.com	216.65.11.111
	www.fbrshop.com	173.254.62.161

6　结束语

本文首先介绍了恶意代码网络行为分析的重要意义；然后结合现有文献提出了恶意代码网络分析的应有内容；再次比较了现有分析系统的差异，选择使用 CUCKOO 系统作为恶意代码网络行为分析的工具；最后对该系统的网络行为分析功能与分析需求进行对比，提出了合理的扩展方案。

参考文献：

[1] BAYER U, MOSER A, KRUEGEL C, *et al*. Dynamic analysis of malicious code[J]. Journal in Computer Virology, 2006, 2(1):66-77.

[2] BAYER U, HABIBI I, BALZAROTTI D, *et al*. A view on current malware behaviors[A]. Proceedings of the 2nd Usenix Workshop on Large-Scale Exploits and Emergent Threats (LEET'09)[C]. Boston, MA, 2009.

[3] 诸葛建伟, 韩心慧, 叶志远等, 僵尸网络的发现与跟踪[A]. 全国网络与信息安全技术研讨会[C]. 2005.183-189.
ZHUGE J W, HAN X H, YE Z Y,*et al*.Discover and track Botnets[A]. NetSec2004[C]. 2005.183-189.

[4] NIVARGI V, BHAOWAL M, LEE T. Machine learning based botnet detection[EB/OL]. http://www,stanford,edu/class/cs229/proj2006/ NivargiBhaowalLee-MachineLear ningBasedBotnetDetection,pdf, 2006-10-10/2008.

[5] KONDO S, SATO N. Botnet traffic detection techniques by C&C session classification using SVM[A]. Proc of the 2nd International Workshop on Security[C]. Berlin: Springer, 2007.91-104.

[6] KUGISAKI Y, KASAHARA Y, HORI Y. Bot detection based on traffic analysis[A]. Proc of 2007 International Conference on Intelligent Pervasive Computing (IPC2007)[C]. Washington, DC, 2007.303-306.

[7] LEE J S, JEONG H C, PARK J H, *et al*. The activity analysis of malicious http-based botnets using degree of periodic repeatability[A]. Proc of 2008 International Conference on Security Technology (SecTech2008)[C]. Washington, DC, 2008. 83-86.

[8] 王威，方滨兴，崔翔. 基于终端行为特征的 IRC 僵尸网络检测[J]. 计算机学报, 2009, 32(10):1980-1988.
WANG W, FANG B X, CUI X. IRC Botnet detection based on host behavior[J].Chinese Journal of Computers, 2009, 32(10):1980-1988.

[9] GU G, PORRAS P, YEGNESWARAN V. BotHunter: detecting malware infection through ida-driven dialog correlation[A]. Proc of the 16th USENIX Security Symp(Security 2007)[C]. 2007.

[10] GU G, ZHANG J, LEE W. BotSniffer: detecting Bomet command and control channels in network traffic[A]. Proc of the 15th Annual Network and Distribut System Security Symp(NDSS'08)[C]. SanDiego, CA, 2008.209-221.

[11] RAMACHANDRAN A, FEAMSTER N, DAGON D. Revealing Botnet membership using DNSBL counter-intelligence[A]. Proc of the Conference on Botnet Detection: Countering the Largest Security Threat[C]. Berlin: Springer, 2008.131-142.

[12] TU H, LI Z T, LIU B. Detecting botnets by analyzing DNS traffic[A]. Proc of the Pacific Asia Workshop on Intelligence and Security Informatics[C]. Berlin: Springer, 2007.323-324.

[13] VILLAMARIN-SALOMON R, BRUSTOLONI J C. Identifying botnets using anomaly detection techniques applied to DNS traffic[A]. Proc of the 5th IEEE Consumer Communications and Networking Conference[C]. Washington, DC, 2008.476-481.

[14] CHOI H, LEE H. Botnet detection by monitoring group activities in DNS traffic[A]. Proc of the 7th IEEE International Conference on Computer and Information Technology[C]. Washington, DC, 2007.715-720.

[15] http://labs,snort,org/papers/zeus,html[EB/OL].

[16] HUANG H D, LEE C S, KAO H Y, *et al*. Malware behavioral analysis system: TWMAN[A]. Intelligent Agent (IA), 2011 IEEE Symposium on[C].2011.11-15.

[17] SONG D, BRUMLEY D, YIN H, *et al*. Bitblaze: a new approach to computer security via binary analysis[A]. Proceedings of the 4th International Conference on Information Systems Security (ICISS'08, keynote invited paper)[C]. Hyderabad, India, 2008.

[18] BELLARD F. A fast and portable dynamic translator[A]. Proceedings of USENIX Annual Technical Conference[C]. USA,2005.41-46.

[19] QEMU: the open source processor emulator[EB/OL]. http://fabrice, bellard,free,fr/qemu/about, htm.

[20] YIN H, SONG D, EGELE M, *et al*. Panorama: capturing systemwide information flow for malware detection and analysis[A]. Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS'07)[C]. New York, NY, USA, 2007.116-127.

[21] AMIT V, Wildcat: an Integrated Stealth Environment for Dynamic Malware Analysis[D]. University of Texas, 2007.

[22] DINABURG A, ROYAL P, SHARIF M, *et al*. Ether: malware analysis via hardware virtualization extensions[A]. Proceedings of the ACM Conference on Computer and Communications Security (CCS'08)[C]. Alexandria, Virginia, USA, 2008. 27-31.

[23] WILLEM S, , CARSTEN H, THORSTE N, *et al*. Toward automated dynamic malware analysis using cwsandbox[A].Proceedings of the IEEE Symposium on Security and Privacy (SSP'07)[C]. 2007.

作者简介：

赵毅（1987-），男，陕西洛南人，东南大学硕士生，主要研究方向为网络安全、恶意代码分析。

龚俭（1957-），男，上海人，东南大学教授、博士生导师，主要研究方向为网络安全、网络行为、网络体系结构。

杨望（1979-），男，安徽宣城人，东南大学讲师，主要研究方向为网络安全、网络管理。