计算机研究与发展
Journal of Computer Research and Development
ISSN 1000-1239/CN 11-1777/TP
47(12): 2037-2048, 2010

僵尸网络检测技术研究进展

王海龙　龚正虎　侯　婕
(国防科学技术大学计算机学院　长沙　410073)
(hlwang@nudt.edu.cn)

Overview of Botnet Detection

Wang Hailong, Gong Zhenghu, and Hou Jie
(*College of Computer, National University of Defense Technology, Changsha* 410073)

Abstract With the rapid development of botnet, the Internet has been facing the growing and disastrous threats. These threats can disable the infrastructure and cause the financial damages, which leads to a severe challenge for the global network security. In order to defense and counter the botnet, the detection is absolutely the basis. Therefore, the research on botnet detection has recently become a hot topic in the field of network security. After analyzing the proposed detection techniques, the authors present the basic process of botnet detection, and make classification for these techniques. Furthermore, according to the different stages of the life cycle of botnet, i.e., propagation, infection, communication and attack, they go into detail about main idea, detection process, merits and shortcomings of the existing techniques. Then, they summarize the approaches and the corresponding algorithms used in the detection techniques, propose the evaluation indices in the six dimensions of source, scope, real-time, accuracy, applicability and flexibility, and compare the representative techniques based on these indices. Later, they discuss the key issues of botnet detection in the fields of multi-source information collection and fusion, essential feature extraction, detection of communication and behavior, correlation analysis and detection architecture. Finally, future research trends are reviewed.

Key words botnet detection; botnet; network security; life cycle; evaluation index

摘　要 僵尸网络的肆虐给互联网带来了极大的威胁,使得僵尸网络检测技术成为近年来网络安全领域的热点研究课题.首先,在对已提出的检测技术进行归纳分析的基础上,概括了僵尸网络检测的基本过程,并对这些检测技术进行了分类;然后,按照僵尸网络生命周期不同阶段的分类方法,着重分析了每种检测技术的研究思路、操作流程和优缺点;接下来,总结了现有检测技术所使用的主要方法及相应算法,提出了评价指标,并对选取的代表性技术进行了比较;最后,探讨了僵尸网络检测的关键问题及今后的研究方向.

关键词 僵尸网络检测;僵尸网络;网络安全;生命周期;评价指标

中图法分类号 TP393

收稿日期:2009-04-14;**修回日期**:2010-04-23

基金项目:国家自然科学基金项目(90604006);国家"八六三"高技术研究发展计划基金项目(2009AA01Z432);国家"九七三"重点基础研究发展计划基金项目(2009CB320503);国家科技支撑计划基金项目(2008BAH37B03)

0 引 言

近年来,分布式拒绝服务(distributed denial of service, DDoS)攻击、垃圾邮件(spam)、网络钓鱼攻击(phishing)、蠕虫(worm)传播、窃取敏感信息等恶意活动已经成为网络安全面临的重要威胁,不仅导致了巨大的经济损失,而且使得全球的网络安全领域面临严峻的考验.究其原因,就是隐秘操纵这些恶意活动的僵尸网络(botnet)[1].

僵尸网络是攻击者(botmaster)利用各种手段传播僵尸程序(bot),将大量主机感染成僵尸主机(zombie),并通过命令与控制(command and control, C&C)信道操纵这些僵尸主机实施恶意行为的网络.它虽然是在网络蠕虫、特洛伊木马、后门工具等传统恶意代码形态的基础上发展、融合而成的一种新型攻击方式[2],但却有着明显的区别:攻击者和僵尸程序之间存在一对多的控制关系[3].正是由于这种关系,僵尸网络比其他恶意程序具备更大的私密性、灵活性和高效性,不仅能够进行可控的主动传播、升级,而且能够通过远程控制从休眠状态迅速转变为攻击状态,造成巨大的破坏.赛门铁克(Symantec)公司的监测数据表明[4-5]:活跃的僵尸主机数从2007年下半年平均每天的61 940台增长为2008年全年平均每天的75 158台,识别出的命令与控制服务器数也从2007年下半年的4 091台激增为2008年全年的15 197台,2008年中国大陆被僵尸网络控制的主机数占全世界总数的比例为13%,再次超过美国成为最大的僵尸网络受害国.

随着蜜罐(honeypot)及蜜网(honeynet)技术的发展[6-7],国内外已经逐步深化了对僵尸网络功能结构、工作原理、命令与控制机制、传播模型等方面的研究.但是蜜罐及蜜网技术存在很多问题[8-9],为了更有效地防御和反制僵尸网络,必须重视和发展检测技术.此外,僵尸网络的固有特点,也对检测技术提出了巨大的挑战:1)僵尸网络行为隐蔽,能够长时间潜伏;2)产生的网络流量很小,能够隐藏在合法的流量中;3)覆盖范围广阔,成员众多,具有一定的组织性;4)僵尸程序能够通过升级增强自身的防御能力,同时改进攻击技术.针对以上问题,国内外学者通过对僵尸网络检测的研究,提出了各种解决方案,取得了许多研究成果.然而,目前还没有详细而全面介绍这些成果的综述论文.为了深入理解僵尸网络检测的机理、存在问题以及发展趋势,并掌握国内外研究的新动向,综述僵尸网络检测技术研究进展工作具有重要的意义.

本文首先阐述了僵尸网络检测的基本思想,对现有检测技术进行分类;接着重点讨论国内外的主要检测技术,并进行了评价和比较;然后分析了僵尸网络检测研究中面临的关键问题;最后总结全文,并展望进一步的研究方向.

1 僵尸网络检测概述

僵尸网络检测技术是随着僵尸网络的发展而发展的.从良性僵尸程序的出现到恶意僵尸程序的实现,从被动传播到利用蠕虫技术主动传播,从使用简单的互联网中继聊天(Internet relay chat,IRC)协议到复杂多变的对等网络(peer to peer,P2P)结构,再到基于超文本传输协议(hypertext transfer protocol,HTTP)及域名系统(domain name system, DNS)协议的控制模式,僵尸网络逐渐发展成规模适中、功能多样、难以检测的恶意网络,给当前的网络安全带来了严重威胁.与此同时,从早期的蜜罐、蜜网的捕获到现在各种检测系统的识别,从针对IRC,HTTP协议到针对P2P,DNS控制,从网络流量分析到恶意行为关联,从垃圾邮件过滤到DDoS防御,检测技术也在不断改进,最终形成种类多样、功能全面、高效准确的网络安全工具,能够及时发现并确认僵尸网络,保证了反制措施的有效实施.

1.1 僵尸网络检测思想

僵尸网络检测的基本思想:首先通过各种途径获取可能存在僵尸网络活动的相关信息,然后根据僵尸网络在这些信息中表征出来的内在特性,应用多种分析技术(统计分析、机器学习、信息理论等)识别并判断出僵尸网络的存在,甚至确定攻击者、命令与控制服务器以及僵尸主机的位置.根据生命周期[9-10]可以把僵尸网络的活动情况分为传播、感染、通信、攻击4个阶段,如图1所示.僵尸网络在每个阶段都有独特的活动模式,这些活动模式会在涉及的网络流量、系统日志、入侵检测系统报告等相关记录中得到体现.对检测出来的僵尸网络信息进行分析,可以进一步获得其他内在特性,既有利于加深对所检测僵尸网络的全面了解,又为以后的分析提供了更多的检测证据.

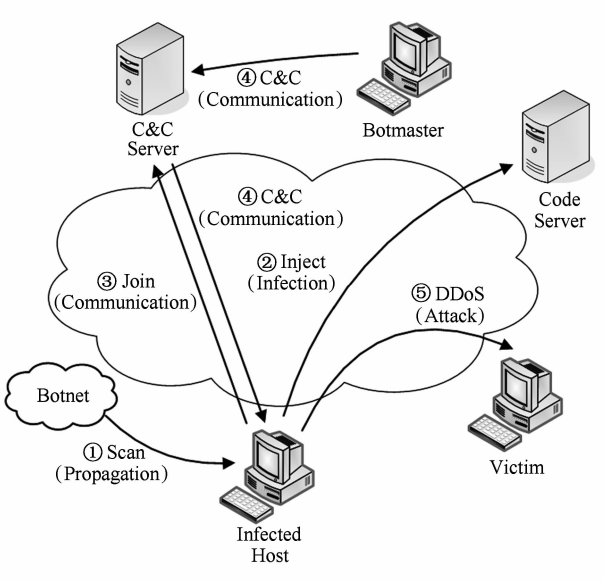

Fig. 1 Life cycle of botnet.

图 1 僵尸网络生命周期

1.2 僵尸网络检测分类

僵尸网络的检测技术根据分类标准的不同,有多种分类方法.文献[11]提出 7 种分类方法:1)基于主机的或基于网络的;2)基于特征码(signature)的或基于行为/异常的;3)检测行为是被动的或主动的;4)检测时机是僵尸程序的传播阶段或执行阶段;5)检测对象是个别的僵尸程序或一组僵尸程序;6)检测过程是否需要附加信息(例如其他系统的报警信息);7)检测技术是约束于或独立于命令与控制技术(例如协议和结构).上述分类从不同侧面初步概括了现有检测技术的差别,但是缺少针对性,大多数分类方法也适合于其他安全检测工具(比如入侵检测系统).因此,针对僵尸网络的活动特点和检测的基本思想,本文提出如下的分类:

根据检测信息的来源分类,可分为蜜罐或蜜网、网络流量数据、日志以及流(flow)数据.

根据检测针对的控制协议分类,可分为 IRC,HTTP,DNS 以及 P2P.

根据检测针对生命周期的不同阶段分类,可分为传播、感染、通信以及攻击.

在生命周期的各个阶段,僵尸网络都会表现出相应的内在特性,不仅有利于分析检测依据的提取,而且可以清晰地展示出检测研究的重点和难点.因此,本文采用第 3 种分类方式选取一些典型的检测方案进行分类讨论.

2 检测技术分析

近年来,学术界提出了许多僵尸网络检测的新技术,其中不少已经在实验环境以及实际网络中得到应用和验证.我们对这些新技术进行研究,选取了部分较为重要的和近期提出的技术,按照生命周期的不同阶段分类,对其核心机制、优缺点等进行了分析和比较.

2.1 传播阶段

僵尸网络主要有 5 种传播形式,包括攻击漏洞、电子邮件携带、恶意网站脚本、即时通信以及伪装软件.除了攻击漏洞,其他 4 种传播形式都必须有用户参与才能成功,存在很大的不确定性.攻击漏洞是僵尸程序能够自主完成的操作,而且通常采用非法扫描获取漏洞,这使得扫描行为成为僵尸网络传播的重要证据.但是由于这种扫描行为隐蔽且容易与其他形式的扫描相混淆,并不易被检测[12].

Li 等人设计了一个能够提取僵尸网络扫描事件的通用模式[13].从蜜网获取的网络流量包含一些带有明显尖峰的稳定背景噪声,这些明显的尖峰通常对应僵尸网络的扫描事件,而且根据流量中背景噪声的尖峰情况能够区分不同的扫描事件.该文设计的模式正是利用了上述特性,首先使用信号分析从流量中分离出扫描事件,然后使用人工分析和可视化技术依次分离出错误配置和蠕虫产生的扫描事件,最终得到僵尸网络的扫描事件.然而,由于很难区分本地扫描是个别扫描行为还是僵尸网络导致的,容易产生误报.而且新感染的僵尸程序不断加入扫描,也很容易误判为蠕虫产生的扫描事件.此外,通过对分辨出来的僵尸网络扫描事件的分析,发现僵尸程序在 IP 分布、自治系统分布、操作系统属性、本地扫描速率等方面都具有一定的特征.

2.2 感染阶段

感染过程是指僵尸网络传播成功后,把僵尸程序植入被攻陷的主机,然后进行更新程序、导入恶意代码、修改 Windows 注册表、关闭特定进程(如防火墙、系统更新)等操作.对僵尸程序而言,感染阶段也应该包括僵尸程序生命周期[10]的一系列行为和活动[14-15].文献[16-18]都是针对感染阶段开展研究工作的.

Liu 等人在文献[16]中指出一个典型的僵尸程序在发作时表现出 3 个不变的特征:1)僵尸程序的启动是自动的,不需要任何用户操作;2)僵尸程序必须和攻击者建立命令与控制信道;3)僵尸程序迟早会执行本地或远程的攻击.该研究小组设计并实现了一个能够在虚拟机技术的帮助下检测上述 3 个特性的系统——BotTracer.它的工作流程主要包括以

下3个阶段:首先,利用白名单Whitelist对所有进程实施过滤,监控剩余的可疑进程;然后,根据提出的命令与控制信道的事件模型,从可疑进程的出入流量中识别出新建立的信道;最后,监控应用程序接口和系统调用,进而确认僵尸程序的存在.实验结果表明BotTracer系统能够成功检测出所有僵尸程序.但是该系统的基本假定是虚拟机不能被僵尸程序检测到,在假定不成立的情况下,检测将会失效.此外,对于在自身启动前就能检测用户行为的僵尸程序,该系统也会失效.

Wurzinger等人提出了一个基于检测模型识别僵尸程序的系统,不依赖任何有关命令与控制信道或传播矢量的信息,也不需要多种感染的关联情况[18].研究发现每个僵尸程序在接收到攻击者的命令后都会以一种特定的方式作出响应行为.正是在这个发现的基础上,该研究小组使用基于特征码和基于异常的方法分别从僵尸程序的命令接收和响应行为两个方面来构建检测模型,并且实现了模型的自动生成.该系统应用于真实的IRC,HTTP以及P2P僵尸程序,都能够自动准确地生成检测模型.实验评估表明,该系统具有较低的误报率,检测性能优于BotHunter系统,但是也存在信道加密及行为模式改变影响检测效果的问题.

2.3 通信阶段

通信过程包括加入僵尸网络过程和控制过程.根据僵尸网络的工作原理,攻击者必须通过命令与控制信道与僵尸主机交互[19].通信是僵尸网络活动必不可少的阶段,而且又要经过网络传输,所以通信阶段是最薄弱的、最易被发现的.命令与控制信道使用的控制协议主要有IRC[20],HTTP[21],DNS[22]以及P2P[23].其中,IRC协议是出现最早、现在仍然应用最广泛的协议,也是被检测的主要对象[24-34];其他3种协议近年来得到了迅速发展,针对它们的研究也引起了普遍关注(HTTP[35-36],DNS[37-42],P2P[43-44]).

2.3.1 针对IRC协议的研究

文献[24-29]中提出的几种比较经典的检测IRC僵尸网络的方法都在文献[2]中作了详细说明.Gu等人提出了一种基于因果关联的主动僵尸网络探测方法,作为对现有被动检测僵尸网络命令与控制技术的有益补充[30].研究表明:1)典型僵尸网络的命令与控制交互具有清晰的命令响应模式,因此无状态的僵尸程序在重复会话中的行为总是确定的,而与人控制的终端交互却是不确定的;2)僵尸程序对接收到的一系列命令的响应行为都是提前编程预置的,和人的反应不同,僵尸程序不能容忍会话中拼写错误的情况.依据这些特性,本文设计了一个算法框架,使用假设检验理论在给定的精度内分辨出僵尸网络的命令与控制会话以及人与人的会话,并在序贯概率比检验(sequential probability ratio test,SPRT)算法的基础上扩展出多种主动探测算法应用在这个框架之中.BotProbe是该框架实现的原型系统,以中间盒的形式插入到网络链路中.通过多种真实的IRC僵尸程序实验验证,BotProbe系统能够成功地识别出确定性的僵尸网络通信,有效地缩短了检测时间,而且在真实用户的参与下也达到了比较低的误报率.但是该主动探测方法只适用于会话式的命令与控制通信的检测,并且在命令与控制信道使用强加密、改变僵尸程序的命令响应时间以及采用有状态的命令与控制协议的情况下,该方法可能会失效.

Nivargi等人提出了基于机器学习的僵尸网络检测技术[31].该技术包括两种方法:一是僵尸网络的二进制检测,不需要任何特征,直接使用机器学习的分类算法(贝叶斯、支持向量机、决策树等)把恶意二进制代码区分出来;二是基于IRC日志的检测,首先从流量中分离出IRC流量,再依据通信特性(信道中的用户数量、IRC命令的数量和频率等),同样应用第1种方法使用的机器学习分类算法从IRC流量中识别出僵尸网络流量.实验结果表明不同的分类算法检测僵尸网络的效果不尽相同,甚至相差很大.

2.3.2 针对HTTP协议的研究

Lee等人在文献[35]中阐明了恶意HTTP僵尸程序是以规律性的间隔反复连接HTTP服务器,不同于正常用户程序.正是根据这个周期性重连的度(degree),第1次提出了一个检测基于HTTP僵尸网络的方法.实验结果表明,检测效果较好,但如果正常用户用程序自动连接HTTP服务器,则可能会产生误报.

Gu等人提出了一个基于网络的异常检测方法,能够识别本地区域网中的僵尸网络命令与控制信道,进而识别出命令与控制服务器以及感染主机[36].文中指出:由于僵尸程序对控制命令的响应行为是编程确定的,相同僵尸网络中的僵尸程序行为具有时空的相关性和相似性;相反,正常的网络行为不太可能表现出同步和相关.根据上述特性,该文针对IRC和HTTP的命令与控制信道设计了BotSniffer系统.该系统使用关联和相似分析的算法检查网络

流量,识别在响应行为上具有很强同步性和关联性的主机群,并把它们作为相同僵尸网络的僵尸程序.通过真实的网络数据评估,BotSniffer 系统能够达到高正确率和低误报率.

2.3.3 针对 DNS 协议的研究

Ramachandran 等人提出了一个被动分析 DNS 黑名单(DNS blacklist,DNSBL)的查询流量来识别僵尸网络的技术[37].文章的主要思想是通过对 DNS 黑名单侦查(reconnaissance)技术的研究,构建一个 DNS 黑名单查询图,而后基于该图分析了合法查询与侦查查询在空间和时间上的关系,并得出第 3 方侦查、自我侦查以及分布式侦查 3 种方式的查询特征.实验结果表明,该技术对早期的僵尸程序检测很有帮助,而且能够进行实时检测,有力支持了对僵尸网络的反制,但在误报率方面需要改进.

Choi 等人提出了一个基于异常的僵尸网络检测机制,通过监控 DNS 流量中的组群(group)行为进行检测[40].该检测机制中包括两个主要算法,分别是僵尸网络 DNS 请求检测算法和僵尸网络迁移检测算法,都是对 DNS 流量中的域名和 IP 地址进行统计分析.在校园网上进行的实验结果表明当僵尸程序连接命令与控制服务器或者迁移到其他服务器上时,该机制都能进行有效检测.然而,如果应用在大规模网络中,算法的处理时间将是检测效率的主要瓶颈.对于有意产生 DNS 请求的欺骗行为,算法就会失去作用.

此外,近年来越来越多的僵尸网络使用快速通量(fast-flux)技术[41],Nazario 等人[42]对快速通量的 DNS 记录进行统计分析,掌握了它的相关特性.该文指出应用快速通量技术的僵尸网络活动通常关联到多个域名,可以在 IP 地址和域名的映射关系的基础上,使用集合论的方法识别这种僵尸网络,并能够根据攻击类别进行分类.

2.3.4 针对 P2P 协议的研究

Schoof 等人在文献[43]中分析了 P2P 僵尸网络,然后针对通信机制提出了几个有利于检测的僵尸网络特性.它们分别是:1)使用开放的连接端口,端口可能是明确的也可能是一个区间;2)NetFlow[45]数据中失败连接的比率非常大;3)使用一个稳定的 IP 地址列表进行对等点发现;4)具有中心点的 P2P 结构(例如 Phatbot[46]).

Steggink 等人[44]以 Peacomm 僵尸网络为例,深入研究了 P2P 僵尸网络,详细描述了如何检测 Peacomm,并通过统计分析提出了能够检测非集中式 P2P 僵尸网络的相关特性.这些特性可以在协议的流量、通过 DNS 查询邮件的交换记录以及 SMTP(Simple Mail Transfer Protocol)连接等统计信息中得到体现.

2.4 攻击阶段

攻击是僵尸网络的最终目的,给整个网络以及用户造成了严重的危害.其中,DDoS 攻击和垃圾邮件[47-48]无疑是危害范围最广、程度最深的,因此它们也受到了专家的格外重视.随着僵尸网络的发展,基于 Web 的攻击带来的危害也逐渐引起了注意.

通过观察发现,DDoS 攻击的本质是许多主机协同、自动的行为,而且攻击者需要一个远程控制大量主机的机制[49].这正好也符合僵尸网络的工作原理.因此,DDoS 攻击的检测方法都可以作为针对某种类型的僵尸网络进行检测的手段.对于 DDoS 攻击检测领域的研究成果[50-51]在这里不作详细介绍.

通过分析垃圾邮件来识别僵尸网络也是比较有效的途径[52-57].Xie 等人提出了一个名为 AutoRE 的自动系统,能够根据统一资源定位符(uniform resource locator,URL)产生检测特征码识别发送垃圾邮件的僵尸网络[52].AutoRE 系统既不需要对输入进行预分类,也不需要训练数据或白名单.利用僵尸网络产生的垃圾邮件通常以一种聚合的模式发送,因此其内容具有相似性.该系统就是基于内容相似性特征,统计得出特征码,输出高质量的正则表达式,使得检测的误报率达到最低.此外,通过对识别出的垃圾邮件活动进行分析,文章揭示了僵尸网络的一些重要特性:1)僵尸主机遍布互联网,从个体角度看发送模式与正常的服务器并无差别;2)僵尸主机发送模式(例如每个电子邮件的收件人数量、连接比率以及发送给无效用户的频率)是可以聚类的,而且发送时间是同步的;3)通过垃圾邮件流量模式的比较,可以得出僵尸网络逐渐成为垃圾邮件的主要发送者,同时躲避检测的能力也在逐步提高.

Duan 等人开发了一个检测发送垃圾邮件僵尸主机的系统——SPOT[53].SPOT 系统是基于序贯概率比检验算法设计的,把网络内的主机是否为僵尸主机作为检验假设,把流出网络的电子邮件消息作为一个事件序列.该系统应用于校园网的大规模电子邮件记录,研究表明 SPOT 系统能够有效的自动检测出网络中的僵尸主机.但也存在以下问题:1)对于批量具有相同或相似内容的垃圾邮件的检测,该方法将会失效;2)系统中垃圾邮件过滤器分离出待检消息的准确性也会影响到整个系统的检测效

果.与 BotHunter 系统不同的是 SPOT 系统是一个轻量级的检测发送垃圾邮件僵尸主机的系统,不需要网络入侵检测系统的支持.

Zhao 等人[54]设计并实现了一个新型的检测僵尸网络实施垃圾邮件攻击的系统——BotGraph.研究发现僵尸程序冒充用户注册和发送电子邮件时会共享 IP 地址.BotGraph 系统正是利用随机图理论(random graph theory)检测这种 IP 地址共享的异常.通过构建巨大的用户连接图并查找紧密的连通子图,BotGraph 系统揭示了僵尸网络活动中的相关性,为识别潜在的僵尸程序冒充的用户提供了有效的检测途径.此外,BotGraph 系统使用新的分布式计算模型在大规模的计算机集群上构建和分析这个巨大的连接图.真实数据实验表明,BotGraph 系统能够从巨大数量的日志中识别出众多僵尸网络产生的用户帐户,而且误报率很低.

基于 Web 的攻击主要包括 SQL(structured query language)注入、代码包含、跨站脚本以及远程文件包含.Robledo 在文献[58]中针对远程文件包含,提出一个检测僵尸网络的方法.该方法的主要思想是通过对 Web 服务器日志中域名、内容以及动态 IP 地址的统计分析,识别出发起攻击的主机.此外,该文还提出了一个跟踪实施远程文件包含攻击的僵尸网络的方法.

2.5 其 他

除了针对生命周期不同阶段的检测技术外,国内外研究者还提出了一些通用的检测框架[59-61].

Gu 等人在文献[14,36]研究基础上,提出了一个通用的检测框架,不依赖僵尸网络命令与控制的协议和结构,不需要任何先验知识(例如捕获的僵尸程序代码、僵尸网络特征码以及命令与控制服务器的名字或地址)[62].该文指出僵尸网络的本质特性是僵尸程序和命令与控制服务器(或对等点)通信并执行恶意行为,而且所做的一切都是以一种相似或相关的方式进行的.因此,该检测框架是对相似的通信和恶意流量进行聚类,然后使用交叉关联的思想(评分函数)识别出具有相似通信和恶意行为模式的主机.BotMiner 是该框架实现的原型系统,主要由通信平面监测(C-plane monitor)、活动平面监测(A-plane monitor)、通信平面聚类(C-plane clustering)、活动平面聚类(A-plane clustering)以及跨平面关联(cross-plane correlation)5 部分组成.其中,通信平面监测模块负责收集和存储僵尸网络中命令与控制通信的流量,活动平面监测模块负责检测可疑活动的流量,通信平面聚类模块和活动平面聚类模块分别对监测模块产生的日志作出聚类处理,跨平面关联模块则结合通信与活动两个平面上聚类模块的结果得出某个主机是否是僵尸网络成员的最终决定.通过对真实网络数据的评估,结果显示 BotMiner 系统能够检测出 IRC,HTTP 以及 P2P 僵尸网络,而且误报率非常低.

3 评价比较

通过第 2 节对各种僵尸网络检测技术的分析,可以看出这些技术都是在观察研究僵尸网络内在特性的基础上,选用恰当的方法及相应的算法,才能充分挖掘可疑信息,进而作出对僵尸网络的准确判断,甚至发现僵尸程序与攻击者的潜在关系.第 2 节中分析的检测技术主要涉及到如下几种方法:图论[37,54]、特征码匹配、统计分析、信号处理[13]、流量(或流等)挖掘、关联分析以及人工分析.特征码匹配方法包括正则表达式[52]、白(或黑)名单[16]、N 元语法(N-gram)模型[14].统计分析方法是指对可疑信息(包括报文大小、电子邮件内容、DNS 查询等)进行方差[18]、比率[43]、概率[30,53]或分布[40,58]等数值计算,判断结果是否满足给定的要求(比如阈值).流量(或流等)挖掘方法是指选取流量(或流等)信息中的某些属性数据,然后使用分类或聚类算法分辨出僵尸网络的活动信息,比如机器学习中的分类算法[31]、X 均值(X-means)聚类算法[62].关联分析方法通常与其他方法结合使用,是依据时空等关系的相关性与相似性对其他方法分析的结果作出综合评估(例如权重分析),包括评分函数[62]、相关矩阵[14]、自相关分析[36].人工分析方法通常是在可视化技术的基础上作出人工判断[13].每种方法及相应的算法都有着自己的优缺点.图论算法可以应用于巨大的数据集分析,它能够帮助发掘大规模僵尸程序活动间的潜在关联性,但必须提供强大的计算资源来保证其在有效时间内获得需要的分析结果.特征码匹配方法虽然能够精确检测,并可以满足实时性或准实时性要求,但是只能检测已知的僵尸网络.统计和信号分析方法能够发现未知的僵尸网络,而且适合于在线和离线的检测,但是容易产生误报,也无法划分不同类型的僵尸网络.流量(或流等)挖掘方法也能够检测未知的僵尸网络,比较适合离线分析,还可以帮助确定攻击者、命令与控制服务器以及僵尸主机的位置,但是具体算法的选择、属性集合的选取以

及训练集的优劣都会影响检测的精度.为了弥补不同方法及相应算法的不足,多种分析手段的集成,尤其是关联分析的使用,能够全面提高僵尸网络检测的效果[62].

对于不同的检测技术可以从以下几个方面进行比较:

1) 来源.指采集被检测的可疑信息的源头,包括蜜罐或蜜网、网络流量数据、日志、流数据.

2) 范围.指检测机制能够实施的区域,包括个人主机、区域网、骨干网.

3) 实时性.指是否能够进行实时或近实时检测.

4) 准确性.由漏报率和误报率来度量,可分为高、中、低.

5) 应用性.由检测僵尸网络类型的数量来度量,可分为强、中、弱.

6) 适应性.由僵尸网络反检测能力对检测技术影响的大小来度量,可分为强、中、弱.

几个典型僵尸网络检测技术的比较情况如表1所示:

Table 1 Comparison of Botnet Detection Techniques

表1 僵尸网络检测技术比较

Technique	Source	Scope	Real-time	Accuracy	Applicability	Flexibility
Ref.[13]	Honeynet	Area	No	Low,High False-positive	High	Medium
BotTracer[16]	Traffic	Host	Yes	Medium	High	Low
BotHunter[14]	Traffic	Area	Yes	Medium	Medium	Medium
Ref.[18]	Honeynet Traffic	Area	Yes	High	High	Medium
BotProbe[30]	Traffic	Area	Yes	Medium	Medium	Low
Ref.[31]	Log	Backbone	No	High	Low,Only IRC Botnet	Medium
BotSniffer[36]	Traffic	Area	No	High	Medium	Medium
Ref.[37]	Log	Backbone	Yes	Low,High False-positive	Low,Only Spam Botnet	Medium
Ref.[40]	Traffic	Backbone	No	Low	Medium	Low
Ref.[43]	Flow	Backbone	Yes	Low	Low,Only P2P Botnet	Medium
AutoRE[52]	Traffic	Backbone	No	High	Low,Only Spam Botnet	Medium
SPOT[53]	Traffic	Area	Yes	High	Low,Only Spam Botnet	Low
BotGraph[54]	Log	Backbone	No	Medium	Low,Only Spam Botnet	Medium
Ref.[58]	Log	Backbone	No	Low	Low,Only HTTP botnet	Medium
BotMiner[62]	Traffic Flow	Area	No	High	High	High

4 关键问题分析

4.1 多源信息采集与融合问题

多源信息采集与融合是有效检测的前提.由于僵尸网络的隐秘性,其踪迹往往隐藏在各种信息之中,分散在不同层次(包括个人主机、区域网、骨干网等)不同类型的网络设备上,以不同的格式进行存储,而且多源信息又包含各种冗余、不确定信息,这些都会对获取可用的检测数据造成影响.

采集与融合是一个有机整体,采集必须满足融合对数据完整性的要求,融合也应该对采集策略起到相应的指导性作用.因此,采集手段必须具有协同性、分布式、智能化的特点,能够根据策略从原始信息中过滤掉无关信息,减少存储空间和传输时间.融合方法必须具有准确性高、复杂度低、扩展性好的特点,先对多源信息提供统一的数据表示和存储,再进行精化处理,最终获取检测信息,并且能够根据信息融合的变化动态调整采集策略.

4.2 僵尸网络内在特性提取问题

内在特性的提取是有效检测的关键.现有检测技术主要采用两种提取方式:1)针对蜜罐或蜜网捕获的僵尸网络样本(包括僵尸程序、通信报文内容等),由于样本数据比较纯净,提取出来的数据特性可以直接作为僵尸网络的内在特性;2)针对通用信息(例如流量数据、日志等),首先要发现信息数据中的特性,然后跟已经论证过的僵尸网络检测系统所发现的结果进行比对,验证信息中的数据特性是否

属于僵尸网络的内在特性.两种方式的关键都需要对复杂数据的概括能力.目前提取过程仍需要安全人员参与,甚至很多有用的信息需要手工分析得到,不能满足实际需求[63].

为了实现提取过程的自动化,同时具备一定的学习能力,将模糊数学、进化计算、机器学习、粗集理论、数据挖掘、人工智能等技术有机的结合起来是一个重要的发展趋势.

4.3 检测僵尸网络通信问题

对通信的检测仍然是检测僵尸网络的主要手段.在通信阶段,重点是针对流量数据或流数据.基于IRC,DNS这种具有明显集中式结构的控制协议,总是表现出比较强的相似、关联等特性,检测效果明显,而且已经提出了很多通用的检测方法.然而,基于HTTP,P2P协议的僵尸网络却表现出较强的个体特性,一直缺乏有效的通用检测手段.特别是随着P2P协议的发展,不同类别的差异性增大,针对僵尸网络应用的反检测能力也得到了提高[64-65].可见,提取基于P2P(或HTTP)协议的僵尸网络在通信上的共同特性是亟待解决的问题.

与区域网和个人主机相比,骨干网具有带宽高、流量大、存储受限以及实验环境要求苛刻等特点,这些因素导致针对僵尸网络通信的实时检测技术发展缓慢,特别是能够应用于核心路由器或交换机的.为了满足检测的实时性要求,必须在充分研究僵尸网络在骨干网中活动规律的基础上,选用实时性较高的方法进行检测.如何能够在保证实时性的前提下提高检测的准确性也将是一个挑战.

4.4 检测僵尸网络行为问题

近几年,由于攻击者反检测能力的提高,僵尸网络通信的不足得到了改进,增大了通信检测的难度[66-67].然而,僵尸程序的行为同样能够表现出其应有的特性.在感染阶段,就是针对僵尸程序的行为模式进行检测的.但是僵尸程序的行为模式又存在以下特点:不同类型的僵尸程序可能有着不同的行为模式;即使是同一个类型,在不同时间或空间上的行为也可能不同.

为了不受僵尸程序类别和行为模式的影响,对于行为本身,除了应该选取更明显的特征外,还必须发现它们在时空属性上的相关性.如果能够把基于主机和基于网络的检测相结合,效果将会更好.

4.5 检测的关联分析问题

通过分析发现,针对某一阶段、单一特性、单一信息源的检测往往不能保证检测的准确性.僵尸网络在生命周期的各个阶段(包括传播、感染、通信、攻击)都表现出协同的本质,这揭示了其具有的时空特性[68].而且在检测的整个过程中,也会出现不同层次上信息相关的情况.忽视上述的关联问题,必然会影响检测的效果.

检测中的关联分析应该包括3个层次:数据、特征、决策.数据关联是在采集的原始信息上进行处理,组合具有明显联系的数据.特征关联是根据检测的目标,将多个特征进行分组,形成统一的特征向量.决策关联是一个高层次的关联,对不同检测模块(或子系统)所得出的结果进行综合分析.根据上述关联思想,构建一个有效的关联分析模型,将成为检测系统的核心模块.

4.6 体系结构问题

现有的僵尸网络检测系统体系结构主要存在以下问题:1)集中式结构不适合大规模网络环境;2)内在特性提取的灵活性不好,大多数结构没有独立的内在特性提取功能,或者提取功能简单,不能集成多种分析方法;3)体系结构中缺少协同功能,虽然实现了分布式采集和监控,但仍需要一个集中的处理中心,缺乏检测系统之间以及与其他安全系统之间有效的信息共享与配合联动;4)协同方式单一,通常只在分析结果上进行一定的信息共享,导致不能对广泛分布的僵尸网络活动作出快速反应.

因此,体系框架必须满足层次化、分布式、可配置、适应性强、扩展性好等要求,能够在检测系统之间以及与其他安全系统之间实现协同.基于主体的检测系统可能是理想的解决方案.

5 总结和展望

僵尸网络的快速发展对互联网安全造成了严重威胁,推动了对僵尸网络检测技术的深入研究.本文首先阐述了僵尸网络检测的基本思想和分类情况;接着按照生命周期的不同阶段分类重点讨论了国内外的主要检测技术,并进行了评价和比较;然后分析了该领域中的关键问题.基于以上分析,下一步研究应重点解决以下问题:1)灵活高效的多源数据采集与融合机制;2)具有学习能力的自动提取僵尸网络内在特性的机制;3)针对P2P(或者HTTP)僵尸网络的通用检测方法以及应用于骨干网的实时检测方法;4)基于行为时空特性的僵尸网络检测机制;5)基于数据、特征、决策的3层关联分析模型;6)具有协同能力的僵尸网络检测系统.

参 考 文 献

[1] Geer D. Malicious bots threaten network security [J]. IEEE Computer, 2005, 38(1): 18-20

[2] Zhuge Jianwei, Han Xinhui, Zhou Yonglin, et al. Research and development of botnets [J]. Journal of Software, 2008, 19(3): 702-715 (in Chinese)
(诸葛建伟,韩心慧,周勇林,等. 僵尸网络研究与进展[J]. 软件学报, 2008, 19(3): 702-715)

[3] Rajab M, Zarfoss J, Monrose F, et al. A multi-faceted approach to understanding the botnet phenomenon [C] //Proc of the 6th ACM SIGCOMM Conf on Internet Measurement Conference(IMC'06). New York: ACM, 2006: 41-52

[4] Dean T, Marc F, Eric J, et al. Symantec global Internet security threat report: Trends for July-December 07(Volume XIII)[R]. Cupertino, CA, USA: Symantec Inc., 2008

[5] Marc F, Eric J, Mack T, et al. Symantec global Internet security threat report: Trends for 2008(Volume XIV)[R]. Cupertino, CA, USA: Symantec Inc, 2009

[6] Baecher P, Koetter M, Holz T, et al. The Nepenthes platform: An efficient approach to collect malware [G] // LNCS 4219: Proc of the Int Symp on Recent Advances in Intrusion Detection (RAID'06). Berlin: Springer, 2006: 165-184

[7] Cheng Jieren, Yin Jianping, Liu Yun, et al. Advances in the honeypot and honeynet technologies [J]. Journal of Computer Research and Development, 2008, 45(Suppl): 375-378 (in Chinese)
(程杰仁,殷建平,刘运,等. 蜜罐及密网技术研究进展[J]. 计算机研究与发展, 2008, 45(增刊): 375-378)

[8] Zhu Zhaosheng, Fu Zhi Judy, Lu Guohan, et al. Botnet research survey [C] //Proc of the 32nd Int Computer Software and Applications Conference. Washington, DC: IEEE Computer Society, 2008: 967-972

[9] Govil J, Govil J. Criminology of botnets and their detection and defense methods [C] //Proc of 2007 IEEE Int Conf on Electro/Information Technology (EIT2007). Washington, DC: IEEE Computer Society, 2007: 215-220

[10] Govil J. Examining the criminology of bot zoo [C] //Proc of the 6th Int Conf on Information, Communications and Signal Processing. Washington, DC: IEEE Computer Society, 2007: 473-478

[11] Gu Guofei. Correlation-based botnet detection in enterprise networks [D]. Atlanta, USA: Georgia Institute of Technology, 2008

[12] Gao Yan, Zhao Yao, Schweller R. Detecting stealthy spreaders using online outdegree histograms [C] //Proc of the 15th IEEE Int Workshop on Quality of Service. Washington, DC: IEEE Computer Society, 2007: 145-153

[13] Li Zhichun, Goyal A, Chen Yan. Honeynet-based botnet scan traffic analysis [C] //Proc of the Conf on Botnet Detection: Countering the Largest Security Threat. Berlin: Springer, 2008: 25-44

[14] Gu Guofei, Porras P, Yegneswaran V, et al. BotHunter: Detecting malware infection through IDS-driven dialog correlation [C] //Proc of the 16th USENIX Security Symp (Security'07). Berkeley, CA: USENIX Association, 2007: 167-182

[15] Stinson E, Mitchell J C. Characterizing bots' remote control behavior [C] //Proc of the 4th Int Conf on Detection of Intrusions and Malware, and Vulnerability Assessment. Washington, DC: IEEE Computer Society, 2007: 89-108

[16] Liu Lei, Chen Songqing, Yan Guanhua, et al. BotTracer: Execution-based bot-like malware detection [G] //LNCS 5222: Proc of the 11th Int Conf on Information Security. Berlin: Springer, 2008: 97-113

[17] Al-Hammadi Y, Aickelin U, Greensmith J. DCA for bot detection [C] //Proc of IEEE Congress on Evolutionary Computation. Washington, DC: IEEE Computer Society, 2008: 1807-1816

[18] Wurzinger P, Bilge L, Holz T, et al. Automatically generating models for botnet detection [G] //LNCS 5789: Proc of the 14th European Symp on Research in Computer Security. Berlin: Springer, 2009: 232-249

[19] Mitsuaki A, Takanori K, Masayoshi S. A proposal of metrics for botnet detection based on its cooperative behavior [C] //Proc of 2007 Int Symp on Applications and the Internet-Workshops. Washington, DC: IEEE Computer Society, 2007: 82-82

[20] Zhuge Jianwei, Holz T, Han Xinhui, et al. Characterizing the IRC-based botnet Phenomenon [R]. Beijing: Peking University & University of Mannheim, 2007

[21] Daswani N, Stoppelman M. The anatomy of clickbot. A [C] //Proc of the 1st Workshop on Hot Topics in Understanding Botnets (HotBots 2007). Berkeley, CA: USENIX Association, 2007: 11-11

[22] Simon H. Working the botnet: How dynamic DNS is revitalising the zombie army [J]. Network Security, 2007, 2007(1): 9-11

[23] Holz T, Steiner M, Dahl F, et al. Measurements and mitigation of peer-to-peer-based botnets: A case study on storm worm [C] //Proc of the 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET'08). Berkeley, CA: USENIX Association, 2008: 73-81

[24] Binkley J R, Singh S. An algorithm for anomaly-based botnet detection [C] //Proc of USENIX SRUTI'06. Berkeley, CA: USENIX Association, 2006: 43-48

[25] Binkley J R. Anomaly-based botnet server detection [C] // Proc of the FloCon 2006 Analysis Workshop. Berkeley, CA: USENIX Association, 2006: 7-12

[26] Strayer W T, Walsh R, Livadas C, et al. Detecting botnets with tight command and control [C] //Proc of the 31st IEEE Conf on Local Computer Networks(LCN'06). Washington, DC: IEEE Computer Society, 2006: 195-202

[27] Livadas C, Walsh R, Lapsley D, et al. Using machine learning techniques to identify botnet traffic [C] //Proc of the 2nd IEEE LCN Workshop on Network Security (WoNS'2006). Washington, DC: IEEE Computer Society, 2006: 967-974

[28] Goebel J, Holz T. Rishi: Identify bot contaminated hosts by IRC nickname evaluation [C] //Proc of USENIX HotBots'07. Berkeley, CA: USENIX Association, 2007: 8-8

[29] Karasaridis A, Rexroad B, Hoeflin D. Wide-scale botnet detection and characterization [C] //Proc of USENIX HotBots'07. Berkeley, CA: USENIX Association, 2007: 7-7

[30] Gu Guofei, Yegneswaran V, Porras P, et al. Active botnet probing to identify obscure command and control channels [C] //Proc of 2009 Annual Computer Security Applications Conf (ACSAC'09). Washington, DC: IEEE Computer Society, 2009: 241-253

[31] Nivargi V, Bhaowal M, Lee T. Machine learning based botnet detection [EB/OL]. (2006-10-10) [2008-12-19]. http://www.stanford.edu/class/cs229/proj2006/NivargiBhaowalLee-MachineLearningBasedBotnetDetection.pdf

[32] Mazzariello C. IRC traffic analysis for botnet detection [C] //Proc of the 4th Int Symp on Information Assurance and Security. Washington, DC: IEEE Computer Society, 2008: 318-323

[33] Kondo S, Sato N. Botnet traffic detection techniques by C&C session classification using SVM [G] //LNCS 4752: Proc of the 2nd Int Workshop on Security. Berlin: Springer, 2007: 91-104

[34] Kugisaki Y, Kasahara Y, Hori Y. Bot detection based on traffic analysis [C] //Proc of 2007 Int Conf on Intelligent Pervasive Computing (IPC2007). Washington, DC: IEEE Computer Society, 2007: 303-306

[35] Lee J S, Jeong H C, Park J H, et al. The activity analysis of malicious http-based botnets using degree of periodic repeatability [C] //Proc of 2008 Int Conf on Security Technology (SecTech2008). Washington, DC: IEEE Computer Society, 2008: 83-86

[36] Gu Guofei, Zhang Junjie, Lee W. BotSniffer: Detecting botnet command and control channels in network traffic [C] //Proc of the 16th Annual Network and Distributed System Security Symposium (NDSS'08). Berkeley, CA: USENIX Association, 2008: 193-210

[37] Ramachandran A, Feamster N, Dagon D. Revealing botnet membership using DNSBL counter-intelligence [C] //Proc of the Conf on botnet Detection: Countering the Largest Security Threat. Berlin: Springer, 2008: 131-142

[38] Tu Hao, Li Zhitang, Liu Bin. Detecting botnets by analyzing DNS traffic [G] //LNCS 4430: Proc of the Pacific Asia Workshop on Intelligence and Security Informatics. Berlin: Springer, 2007: 323-324

[39] Villamarin-Salomon R, Brustoloni J C. Identifying botnets using anomaly detection techniques applied to DNS traffic [C] //Proc of the 5th IEEE Consumer Communications and Networking Conf. Washington, DC: IEEE Computer Society, 2008: 476-481

[40] Choi H, Lee H, Lee H. Botnet detection by monitoring group activities in DNS traffic [C] //Proc of the 7th IEEE Int Conf on Computer and Information Technology. Washington, DC: IEEE Computer Society, 2007: 715-720

[41] Holz T, Gorecki C, Rieck K, et al. Measuring and detecting fast-flux service networks [C] //Proc of the 16th Annual Network and Distributed System Security Symp(NDSS'08). Berkeley, CA: USENIX Association, 2008: 181-192

[42] Nazario J, Holz T. As the Net Churns: Fast-flux botnet observations [C] //Proc of the 3rd Int Conf on Malicious and Unwanted Software(Malware2008). Washington, DC: IEEE Computer Society, 2008: 24-31

[43] Schoof R, Koning R. Detecting peer-to-peer botnets [R]. Amsterdam, Holland: University of Amsterdam, 2007

[44] Steggink M, Idziejczak I. Detection of peer-to-peer botnets [D]. Amsterdam, Holland: University of Amsterdam, 2008

[45] Claise B, Sadasivan G, Valluri V, et al. Cisco Systems NetFlow Services Export Version 9 (RFC3954) [S]. Strasbourg, France: Internet Engineering Task Force (IETF), 2004

[46] Grizzard J B, Sharma V, Nunnery C, et al. Peer-to-peer botnets: Overview and case study [C] //Proc of USENIX HotBots'07. Berkeley, CA: USENIX Association, 2007: 1-1

[47] Chiang K, Lloyd L. A case study of the restock rootkit and spam bot [C] //Proc of USENIX HotBots'07. Berkeley, CA: USENIX Association, 2007: 10-10

[48] Ramachandran A, Feamster N. Understanding the network-level behavior of spammers [C] //Proc of ACM SIGCOMM 2006 Conf. New York: ACM, 2006: 291-302

[49] Freiling F, Holz T, Wicherski G. Botnet Tracking: Exploring a root-cause methodology to prevent denial of service attacks[G] //LNCS 3679: Proc of the 10th European Symp on Research in Computer Security. Berlin: Springer, 2005: 319-335

[50] Carl G, Kesidis G, Brooks R R, et al. Denial-of-service attack-detection techniques [J]. IEEE Internet Computing, 2006, 10(1): 82-89

[51] Yan Fen, Wang Jiajia, Zhao Jinfeng, et al. Survey of detection on DDoS attack [J]. Application Research of Computers, 2008, 25(4): 966-969 (in Chinese)
(严芬,王佳佳,赵金凤,等. DDoS攻击检测综述[J]. 计算机应用研究, 2008, 25(4): 966-969)

[52] Xie Yinglian, Yu Fang, Achan K. Spamming botnets signatures and characteristics [J]. Computer Communication Review, 2008, 38(4): 171-182

[53] Duan Zhenhai, Chen Peng, Fernando S, et al. Detecting spam zombies by monitoring outgoing messages [C] //Proc of IEEE INFOCOM 2009. Washington, DC: IEEE Computer Society, 2009: 1764-1772

[54] Zhao Yao, Xie Yinglian, Yu Fang, et al. Botgraph: large scale spamming botnet detection [C] //Proc of the 6th USENIX Symp on Networked Systems Design and Implementation (NSDI'09). Berkeley, CA: USENIX Association, 2009: 321-334

[55] Zhuang Li, Dunagan J, Simon D R, et al. Characterizing botnets from email spam records [C] //Proc of the 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats(LEET'08). Berkeley, CA: USENIX Association, 2008: 10-18

[56] Ian C, Eimear B. The automatic discovery, identification and measurement of botnets [C] //Proc of the 1st Int Workshop on Dependability and Security in Complex and Critical Information System. Washington, DC: IEEE Computer Society, 2008: 127-132

[57] Husna H, Phithakkitnukoon S, Dantu R. Traffic shaping of spam botnets [C] //Proc of the 5th IEEE Consumer Communications and Networking Conf. Washington, DC: IEEE Computer Society, 2008: 786-787

[58] Robledo H F G. Types of hosts on a remote file inclusion (RFI) botnet [C] //Proc of the 5th Electronics, Robotics and Automotive Mechanics Conf. Washington, DC: IEEE Computer Society, 2008: 105-109

[59] Ji S G, Im C T, Kim M J, et al. Botnet detection and response architecture for offering secure Internet services [C] //Proc of 2008 Int Conf on Security Technology (SecTech2008). Washington, DC: IEEE Computer Society, 2008: 101-104

[60] Paxton N, Ahn G J, Chu B. Towards practical framework for collecting and analyzing network-centric attacks [C] // Proc of IEEE Int Conf on Information Reuse and Integration. Washington, DC: IEEE Computer Society, 2007: 73-78

[61] Lu Wei, Ghorbani A A. Bots behaviors vs. human behaviors on large-scale communication networks [G] //LNCS 5230: Proc of the 11th Int Symp on Recent Advances in Intrusion Detection. Berlin: Springer, 2008: 415-416

[62] Gu Guofei, Zhang Junjie, Perdisci R, et al. Botminer: Clustering analysis of network traffic for protocol- and structure-independent botnet detection [C] //Proc of the 17th USENIX Security Symposium(Security'08). Berkeley, CA: USENIX Association, 2008: 139-154

[63] Tang Yong, Luo Jiaqing, Xiao Bin, et al. Concept, characteristics and defending mechanism of worms [J]. IEICE Transaction, 2009, 92-D(5): 799-809

[64] Dong Dafan, Wu Ying, He Liang. Deep analysis of intending peer-to-peer botnet [C] //Proc of the 7th Int Conf on Grid and Cooperative Computing. Washington, DC: IEEE Computer Society, 2008: 407-411

[65] Hund R, Hamann M, Holz T. Towards next-generation botnets [C] //Proc of the 4th European Conf on Computer Network Defence (EC2ND). Washington, DC: IEEE Computer Society, 2008: 33-40

[66] Shirley B, Mano C D. A model for covert botnet communication in a private subnet[G] //LNCS 4982: Proc of the 7th Int IFIP-TC6 Networking Conf. Berlin: Springer, 2008: 624-632

[67] Zou C C, Cunningham R. Honeypot-aware advanced botnet construction and maintenance [C] //Proc of Int Conf on Dependable Systems and Networks. Washington, DC: IEEE Computer Society, 2006: 199-208

[68] Zhang Zonghua, Youki K. A holistic perspective on understanding and breaking botnets: Challenges and countermeasures [J]. Journal of the National Institute of Information and Communications Technology, 2008, 55(2): 43-59

Wang Hailong, born in 1981. Received his MSc degree in computer science from the National University of Defense Technology in 2006. Since 2007, he has been a PhD candidate in computer science of the National University of Defense Technology. He is a student member of the China Computer Federation. His current research interests include network and information security, distributed computing, and computer network architecture.

王海龙, 1981 年生, 博士研究生, 中国计算机学会学生会员, 主要研究方向为网络与信息安全、分布式计算、计算机网络体系结构.

Gong Zhenghu, born in 1945. Professor and PhD supervisor in computer science of the National University of Defense Technology. His main research interests include computer network and communication, network security.

龚正虎, 1945 年生, 教授, 博士生导师, 主要研究方向为计算机网络与通信、网络安全等.

Hou Jie, born in 1983. Received her MSc degree in computer science from the National University of Defense Technology in 2006. Since 2007, she has been a PhD candidate in computer science of the National University of Defense Technology. She is a student member of the IEEE. Her current research interests include computer network architecture, network and information security.

侯　婕, 1983 年生, 博士研究生, IEEE 学生会员, 主要研究方向为计算机网络体系结构、网络与信息安全.

Research Background

Botnet is a new type of attack which is developed and syncretized from network worm, Trojan, backdoor tools and other traditional forms of malicious code. It has been a serious threat to Internet security over ten years, especially in China mainland. With the evolution of botnet, the detection techniques for it have also developed. Many diverse schemes for botnet detection have been proposed, such as honeypot or honeynet for capture and analysis, correlation analysis of malicious behaviors, detection approaches for different C&C mechanisms (e. g. IRC, HTTP, DNS, or P2P), and identifying bots from DDoS and spam. These researches not only play a guiding role in the academic, but in practice have achieved good results (e. g. BotHunter). However, there is no material to give an overview of these achievements. In this paper, we are concentrating on the botnet detection techniques, expounding the basic process, making classification, analyzing their characteristics by category, carrying out the evaluation and comparison, and finally discussing the key issues and development trends. With the overview of botnet detection, we believe that it should be an effective way to identify the bots, even the botmasters, and provide a strong support for botnet tracking, prevention and counter-measure, thus to relieve or eliminate the harm of botnet.

Our work is supported by the National Natural Science Foundation of China under grant No. 90604006, the National 863 High-Tech Research and Development Plan of China under grant No. 2009AA01Z432, the National 973 Basic Research Program of China under grant No. 2009CB320503, and the National Science and Technology Support Program of China under grant No. 2008BAH37B03.