第二篇：应对恶意代码安全威胁（标题）

引言

恶意代码已经成为互联网最严重的安全威胁，据 2006 年美国 CSI 和 FBI 的调查报告，从 2000 年至 2006 年，恶意代码一直是造成受调查单位最大损失的攻击类型，我国公安部组织的 2007 年全国信息网络安全状况调查也显示，国内 91.4%的被调查单位在去年期间感染了计算机病毒或木马等恶意代码，达到历年最高。

作为国内信息安全领域一家知名的科研机构，北京大学计算机所信息安全工程研究中心（下简称工程研究中心）也一直将恶意代码分析、检测及防范技术作为其重要的科研方向之一。工程研究中心以“技术顶天，市场立地”为发展宗旨，坚持基础研究与应用开发并重，与方正集团等企业紧密合作，建立了产学研相结合的信息安全产业链。在恶意代码方向上，工程研究中心承担了多项国家科研项目，积累了丰富的技术经验，技术成果也已经在 CNCERT/CC 等国家相关部门得到实际应用，同时也研发了具有自主知识产权的防虫墙等产品，并已通过方正集团投放国内市场。

以下将从样本采集、样本分析、检测与防范这三个恶意代码处置流程中的关键环节来介绍相关的技术方法，以及工程研究中心所开展的研发工作和成果。

基于蜜罐技术自动采集恶意代码样本

为了有效应对恶意代码所带来的安全威胁，尤其是具有快速传播能力的主动传播型恶意代码，反病毒厂商、计算机安全应急响应部门和安全研究人员都必须在恶意代码传播早期尽早地获取到恶意代码样本，通过进一步深入分析，提取精确的检测特征码以及有效的应对措施，从而及时地对其进行抑制。

传统的恶意代码样本采集方式包括现场提取、用户上报、厂商样本交换等，但这些传统方式一般需要人工参与，但随着恶意代码变种数量近年来的飞速增长，以及恶意代码传播速度的加快（如 Slammer 蠕虫在十分钟内感染了互联网上所有的易感主机），这些需要人工参与的传统恶意代码采集方式无法适应及时应急响应的需求，因此我们需要完全自动化的恶意代码样本采集方式。

工程研究中心狩猎女神项目组在多年的蜜罐技术研究基础上，已构建了结合低交互式蜜罐和高交互式蜜罐技术的全自动化恶意代码样本采集系统，并协助 CNCERT/CC 在全国范围进行了大规模的实际部署和应用，取得了良好的恶意代码监测效果。

低交互式蜜罐－Nepenthes

Nepenthes 是一款著名的基于低交互式蜜罐技术的恶意代码样本自动采集软件，由 Mwcollect.org 团队的 Paul Baecher 和 Markus Koetter 等人开发并开源发布。Nepenthes 软件的基本设计思想是通过模拟存有漏洞的网络服务，构建低交互式的蜜罐系统，使其能够与网

络上传播的恶意代码进行一定程度上的交互，并从交互过程中分析获取恶意代码样本的感染源位置信息，从而对这些恶意代码样本进行自动化地采集。

基于这样的设计思想，Nepenthes 软件实现为 Linux 平台下的守护进程，Nepenthes Core 负责通过网络接口与恶意代码感染源进行交互，并协调其他各类组件模块共同完成恶意代码样本采集任务，目前 Nepenthes 软件中包含如下几类组件模块：①漏洞模拟模块：模拟各种已知的网络服务中存在的安全漏洞，如影响广泛的 LSASS 漏洞（MS04-011）、RPC-DCOM 漏洞（MS03-026）、ASN1 漏洞（MS04-007）等；②Shellcode 分析模块：分析由漏洞模拟模块所接收的攻击负载（Payload），并从中提取恶意代码感染源的位置信息；③获取模块：实现各种协议的恶意代码样本获取功能，通过提取的 URL 位置信息从远程位置下载恶意代码样本；④提交模块：实现各种形式的恶意代码样本提交功能，包括写入本地文件系统，写入数据库，提交远程收集服务器，以及提交给反病毒厂商进行扫描和分析等；⑤日志模块：记录恶意代码捕获过程中的信息，为进一步的恶意代码传播模式分析提供基础数据。⑥其他模块：包括地理位置查询模块和端口监听模块等。

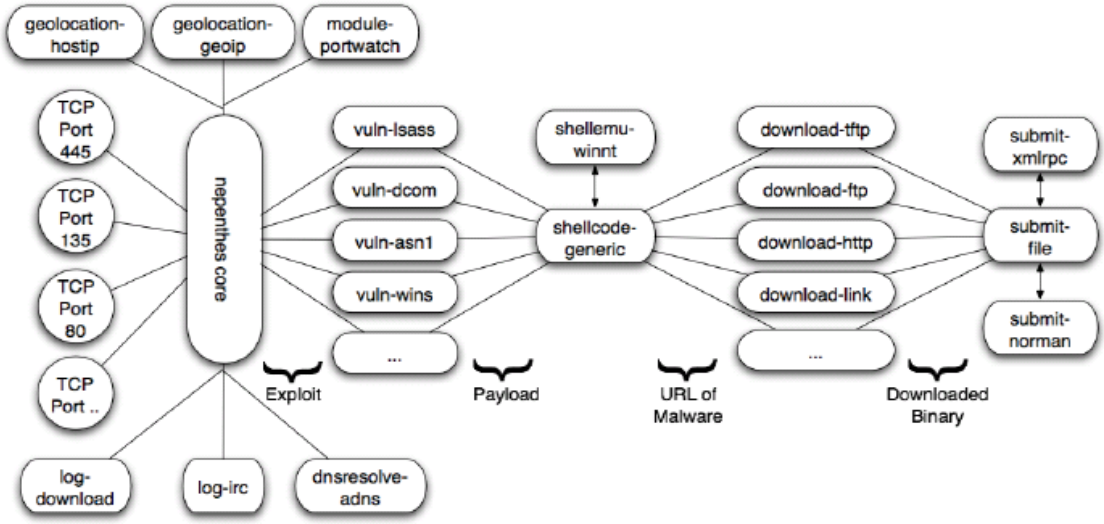

图 1 Nepenthes 软件的组成结构框架

高交互式蜜罐－HoneyBow

HoneyBow 是一款基于高交互式蜜罐技术的恶意代码样本自动采集软件，该软件由北京大学狩猎女神项目组研发，与 Nepenthes 软件一样以 Mwcollect.org 团队的名义开源发布。

HoneyBow 软件集成了三个恶意代码采集工具 MwWatcher、MwFetcher 和 MwHunter，分别采用了不同的技术和策略来检测和采集恶意代码样本，从而达到更为全面的恶意代码捕获效果，特别是针对攻击未知安全漏洞的“零日”恶意代码。MwWatcher 工具基于蜜罐系统的本质特性——“蜜罐系统没有任何的业务活动，因此蜜罐系统中发生的任何行为都预示着恶意”，通过实时监控蜜罐系统文件系统的变化，从而发现和捕获恶意代码样本。MwWatcher 在高交互式蜜罐系统中安装并运行，当主动传播型恶意代码攻击高交互式蜜罐系统上存有安全漏洞的网络服务并试图感染系统时，恶意代码样本将会被传播并保存在蜜罐主机文件系统中，MwWatcher 将通过截获文件系统调用，实时捕获恶意代码样本文件；MwFetcher 工具则通过交叉对比受感染的蜜罐文件系统列表和之前保存的干净文件系统列表间的差异，提取可疑的恶意代码样本，通过差异比对 MwFetcher 能够完备地获取这段时间内感染蜜罐文件系统的恶意代码，包括对高层文件系统 API 调用隐藏的 Rootkit；MwHunter 工具实现为著名开源网络入侵检测系统 Snort 的一个动态处理插件，可以与标准的第三代蜜

网中在蜜网网关上以 inline 模式运行的 Snort 组合使用，MwHunter 依赖于 Snort 的 Stream 处理插件，从经过会话重组的网络流中识别 Windows 平台上的 PE 格式可执行文件，并将文件内容通过启发式方法进行提取和采集。

与基于低交互式蜜罐的 Nepenthes 恶意代码自动采集器相比，HoneyBow 使用了真实的存有安全漏洞的服务来诱骗恶意代码感染，而模拟实现网络服务的方法，因此 HoneyBow 具有自动采集“零日”恶意代码的能力，即使它们所攻击的安全漏洞在爆发前对于整个业界还是未知的。另外，我们也无需深入调查安全漏洞的详细细节，并实现一个模拟版本，因此 HoneyBow 的部署更加灵活和方便。在另一方面，HoneyBow 在可扩展性上较低交互式蜜罐相比具有其局限性，因此在我们的分布式蜜网部署中，我们结合了 HoneyBow 和基于低交互式蜜罐的 Nepenthes 以构建一个更加完整的恶意代码自动采集系统。

Matrix 中国分布式蜜网系统

为了加强对恶意代码的样本采集和监测处理能力，工程研究中心狩猎女神项目组协助 CNCERT/CC 于 2006 年开始陆续在国内 15 个省市部署了 Matrix 中国分布式蜜网系统，并于 2007 年在 11 个省市进行了正式的工程化部署。通过 Matrix 分布式蜜网系统捕获的恶意代码样本，可以掌握我国互联网上主动传播型恶意代码的扩散和流行情况。

2007 年上半年期间，Matrix 分布式蜜网系统每日平均捕获恶意代码样本 3041 次，而消除重复捕获情况，每日捕获的不重复新样本数量达到 442 个，这也说明了目前互联网上的主动传播型恶意代码的泛滥程度。在共计 55 多万次的恶意代码样本捕获中，位于前十位的恶意代码如表所示：

排名	恶意代码名称	总捕获次数
1	Backdoor.Win32.PoeBot.c	62203
2	Backdoor.Win32.VanBot.ax	61622
3	Net-Worm.Win32.Allaple.b	42755
4	Virus.Win32.Virut.b	30964
5	Backdoor.Win32.SdBot.aad	17958
6	Backdoor.Win32.SdBot.xd	16156
7	Backdoor.Win32.Rbot.gen	15030
8	Virus.Win32.Virut.a	14236
9	Net-Worm.Win32.Allaple.e	14070
10	Backdoor.Win32.IRCBot.ul	13843

由于 2007 年上半年并没有大规模传播的蠕虫出现，Matrix 分布式蜜网系统捕获最多的是僵尸程序，它们主要利用微软系统的漏洞进行传播，并在感染的机器上留下后门程序，通过 IRC、HTTP 等协议进行远程控制形成僵尸网络。黑客利用僵尸网络能够窃取被感染主机的系统信息，并控制被感染的机器发起新的扫描、DDoS 攻击、发送垃圾邮件或进行远程控制和网络欺诈活动。如：PoeBot（派波）、VanBot、SDBot、Rbot（瑞波）等僵尸程序均具有较高的危害性。

图 2 Matrix 分布式蜜网恶意代码样本捕获趋势图

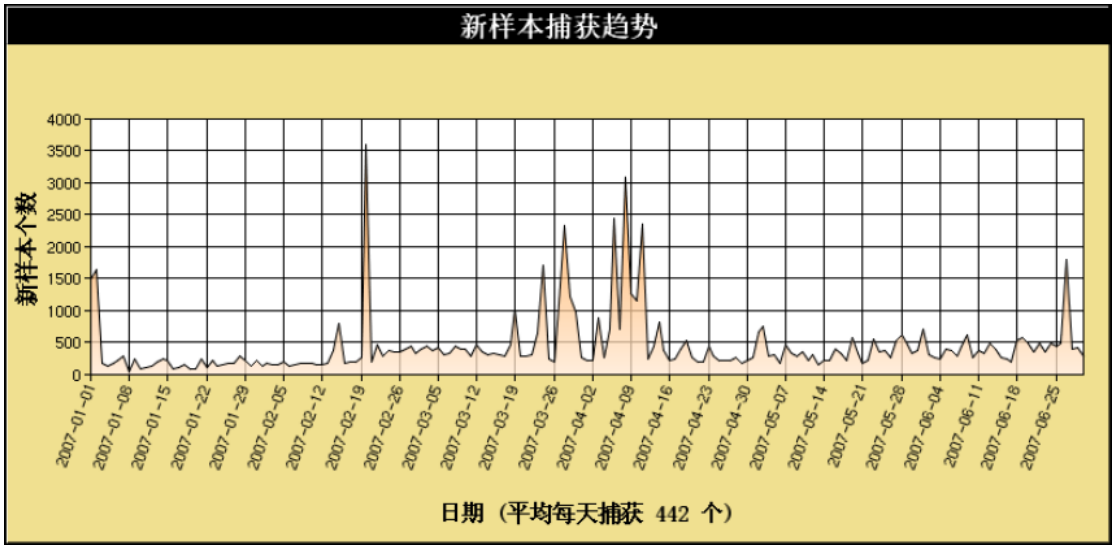

图 3 Matrix 分布式蜜网恶意代码新样本捕获趋势图

恶意代码自动分析技术

在采集获得流行恶意代码样本基础上，进一步的恶意代码分析是对抗恶意代码的核心技术环节，也是信息安全领域所关注的一个重点问题。恶意代码分析的主要目标包括：

1. 识别恶意代码的目标平台、类型及家族；
2. 理解恶意代码的行为及意图，给出恶意代码机理分析报告；
3. 提取恶意代码特征码，以支持对该恶意代码的及时检测与查杀。

构建恶意代码自动分析平台，对大多数恶意代码实行自动化分析，仅将少数无法自动化处理的恶意代码交付给人工做进一步分析，是恶意代码分析技术的发展趋势。

工程研究中心狩猎女神项目组在国家 242 信息安全计划的资助下，目前正在开展恶意代码自动分析平台的研发工作，并已取得了良好的进展。

静态分析

静态分析方法通过对恶意代码二进制程序的逆向工程分析，进行控制流图生成、高层结构恢复等反编译或类反编译（类反编译是指由二进制代码生成某种特定中间语言的过程）工作，并以此为基础分析代码的特定属性。

在静态分析技术方面，目前狩猎女神项目组正在进行以下几方面的研发工作：

1. 实现了集成多个国内外主流反病毒引擎的恶意代码扫描软件 MwScanner，该软件可调用各厂商提供的反病毒引擎，并使用其最新的病毒特征库，对待分析样本进行扫描，从而识别已知恶意代码的家族和分类，并记录样本是否各反病毒引擎所未知的“零日”恶意代码。基于该恶意代码扫描软件，狩猎女神项目组将协助 CNCERT/CC 构建类似于 VirusTotal 的公众恶意代码鉴别服务。
2. 针对目前恶意代码流行加壳的趋势，研究针对流行的恶意代码加壳的识别技术，同时构建基于虚拟执行的恶意代码自动脱壳技术，从而能够有效地脱去恶意代码各式各样的“马甲”，为准确识别和进一步分析恶意代码样本提供基础。
3. 研究恶意代码控制流图、API 调用序列等代码高层结构属性的重构和分析技术，并在此基础上理解恶意代码的行为，以及支持对未知恶意代码的相似性聚类和识别。

动态分析

动态分析方法则通过在受控环境中执行待分析恶意代码，以获取目标代码的行为及运行结果，常用的动态分析方法包括：①系统监控方法，②虚拟环境方法和③动态调试方法。

在动态分析方法方面，目前狩猎女神项目组正在深入研究轻量级沙箱技术，为大规模并行分析的恶意代码样本提供一个高度受控、资源隔离的运行环境，并通过内核态 API 劫持技术监控恶意代码在运行时刻所表现出的行为特征。在轻量级沙箱这一基础技术的协助下，我们就能够高效地分析由 Matrix 分布式蜜网系统捕获的大量恶意代码样本，对其进行自动化分析，帮助 CNCERT/CC 等计算机安全应急响应部门充分了解最新流行的恶意代码行为，并提取其连接的僵尸网络控制信道、恶意代码感染源等关键信息，支持进一步的安全威胁追踪和处置。

特征码提取与验证

为支持在客户端对恶意代码样本进行检测，以及在网络链路上更大范围地对恶意代码传播进行监测，狩猎女神项目组目前也在开展恶意代码特征码的自动提取与验证技术。

一个良好的恶意代码特征码应具备极低误报率、高检出率、鲁棒性等特性，狩猎女神项目组采用了基于行为特异性和结构复杂度的特征码自动化提取技术，使得获取的特征码具有较高的质量，而进一步对特征码进行验证是保证其不会误杀正常文件的关键步骤。前不久诺顿误杀事件的根结也就出在特征码的验证环节，诺顿防病毒软件之所以将简体中文版本 WinXP SP2 系统文件误报为病毒，其根本原因是因为没有将中文版的 Window XP SP2 操作系统文件加入误报库，或者没有同步更新由于 Windows 补丁自动分发和修补所改变的系统文件。

恶意代码的检测与防范

恶意代码检测与防范技术的发展

随着二十世纪八十年代计算机病毒的出现，逐渐兴起了计算机反病毒产业，这些反病毒产品主要以软件的形式出现，通过在计算机内部执行的方式，来清除计算机感染的病毒，这是最早的恶意代码检测与防范系统。

随着互联网技术的蓬勃发展，计算机网络逐步深入到社会各个层次，大量的企事业单位和团体接入到了网络中，病毒的传播途径也迅速拓宽，网络成为了病毒传播的高速公路，网络病毒，已经发展成为包含传统病毒、网络蠕虫、木马、间谍程序等多种形态的恶意代码，它们之间互相融合，在传播能力、生存能力、杀伤性方面均大大提高，是互联网上的主要危害之一。在这样的情况下，传统反病毒技术与网络安全技术形成了一个新的汇合点，一些厂商研发出了在网关进行恶意代码扫描的产品，通过还原网络流中的原始数据，对其进行恶意代码扫描，从而切断恶意代码的网络传输通道，防止其跨网关传染。上述两种防范方法在恶意代码检测的基础技术方面并没有改变，继续沿用了传统的反病毒引擎，这种引擎采用的是特征码技术和虚拟机技术的结合，其中特征码技术是主导。

用特征码检测恶意代码的方案具有一定的局限性。该技术要求从病毒体中提取病毒特征值，所以只有等到新病毒出现后，才有可能获得病毒体，并针对它进行单独处理。这种方法的固有缺陷是，对新病毒的防范始终滞后于病毒的出现。

虚拟机技术是一种启发式探测未知病毒的反病毒技术，目前已经在反病毒软件中得到广泛应用。但由于 PC 的计算能力有限，反病毒软件的制造成本也有限，而病毒的发展可以说是无限的，让虚拟机技术获得更加实际的功效，甚至要以此为基础来清除未知病毒，其难度相当大。而特征码方法是适用范围最宽、速度最快、最简单、最有效的方法，因此在传统反病毒技术中，特征码技术依然是中坚力量。受病毒在理论上就是不可判定的这一根本前提的制约，事实上，无论是启发式，亦或是虚拟机，都只能是一种工程学的努力，针对病毒永远不可达到 100％的检测和清除。

工程研究中心的相关研发进展

僵尸程序、间谍软件、网页木马等网络病毒的频频爆发，已经使国内外反病毒领域开始意识到，单纯依靠 “特征码技术”已经不能适应反病毒需求。能否率先掌握主动防御技术，高效主动防御未知病毒已成为国内外反病毒研究机构和厂商亟待突破的重点。

北京大学计算机所信息安全工程研究中心提出的基于行为针对网络恶意代码进行检测、抑制手段，开辟了一套新的思路，省去了传统病毒防范中获取恶意代码样本，并对其分析、提取特征码的周期，能够在第一时间发现网络病毒的活动，从而采取有效措施对其进行抑制，实现 0-day 的网络病毒防护。

方正防虫墙就是工程研究中心采用行为分析方法，并综合传统恶意代码检测技术，研制的一款恶意代码检测产品，该产品在能够网络层次防范恶意代码入侵，切断恶意代码传播途径。结合主动抑制和被动检测技术，在网络边界、内网等多个层面防范蠕虫、木马和病毒等恶意代码的传播。系统中采取了积极主动的措施，自动对异常网络节点进行隔离，吸收、引导其恶意网络流，保护网络上的其他节点免受侵害，维持整个网络的健康运行。

此外工程研究中心也承担了电子产业基金项目的研究任务，构建了一套蠕虫病毒分布式监控和防范系统，该系统在网络边界和内网设置多道防线，采用流式扫描引擎、网关过滤、行为统计分析等技术，在网络层面防范蠕虫病毒入侵，并切断蠕虫病毒在内网的传播和蔓延。该系统由病毒过滤网关、子网探测器、集中管理监控平台三类设备构成，形成一个立体防护模型。该项目研发的“基于行为的网络蠕虫和恶意代码识别技术”可以有效识别未知蠕虫和网络恶意代码，运用了先进的行为识别和行为学习模型技术，对异常网络的敏感度以及识别的正确性有了进一步的提高；“基于隔离的网络免疫技术”在不改变网络物理拓扑的情况下，对可疑主机进行封禁，并且降低其对网络干扰的一项技术。依据 TCP/IP 协议原理，将蠕虫机的流量诱导至探测设备的吸收端口，同时减少蠕虫感染机在子网内的广播数据包，并切断其与其他网络主机的联接，阻止它继续感染其他机器。

5. 结束语

恶意代码已成为网络攻防的主战场，黑客们已发展出各式各样不同形态的恶意代码以帮助他们自动化地达成攻击目的，因此恶意代码形态、家族和变种层出不穷，也令反病毒研究机构和厂商疲于应付。反病毒研究机构和厂商需要研发更加先进的恶意代码采集、分析、检测与防范技术，才能够在这场旷日持久的技术博弈中占据优势地位。北京大学计算机所信息安全工程研究中心以恶意代码检测与防范技术为其重要的科研方向之一，已经在此方向积累了坚实的科研基础，也愿与国内外业界同仁携手面对来自黑客界的挑战。

恶意代码传播机制

恶意代码按照其传播机制的不同可以分为主动传播型和被动传播型两大类：主动传播型恶意代码包括计算机病毒、蠕虫、僵尸程序等，主要通过文件系统扫描、网络扫描等方式寻找感染目标，并通过攻击服务漏洞、共享文件目录等途径完成其传播。被动传播型恶意代码包括特洛伊木马、网页病毒、邮件病毒等，一般需要用户参与，通过社会工程学方法诱使用户点击或者运行，从而帮助这类恶意代码完成其扩散过程。

恶意代码处置流程

一般反病毒厂商对恶意代码的处置流程包括样本采集、样本分析、特征码提取和升级、以及客户端查杀等主要环节。恶意代码在互联网上爆发后，反病毒厂商首先通过各种样本采集渠道获取恶意代码样本；然后在一个隔离的病毒实验室对样本进行分析，以深入剖析其行为机理和特征，分析结果一般会以恶意代码样本分析报告的方式在互联网上进行发布；在样本分析基础上，反病毒厂商会进一步从恶意代码样本中提取病毒检测特征码，在进行充分的测试后更新其病毒特征库；反病毒厂商客户端的防病毒软件将周期性更新病毒特征库，从而能够对新爆发的恶意代码具备查杀能力。在这样的恶意代码处置流程中，从一个未知恶意代码现身互联网至反病毒厂商的防病毒软件具备对该恶意代码的查杀能力所需时间被称为反病毒厂商对于该恶意代码段的响应时间，反病毒厂商处置恶意代码的平均响应时间是衡量其技术实力的一个重要性指标，而样本采集、样本分析和特征码提取与升级等各个环节的效率都最终决定了这一指标，也决定了反病毒厂商对快速传播恶意代码的应对效果。