面向欺骗防御的蜜网技术研究

刘亚群　高雅卓　邢长友*　张国敏
（中国人民解放军陆军工程大学　江苏 南京 210007）

摘　要　蜜网通过构建诱捕环境并伪装成真实的业务网络来欺骗攻击者，吸引攻击者攻击，监控攻击者的行为并分析其特征，已经成为网络欺骗防御的核心手段。介绍蜜网的定义、分类与功能，在此基础上结合蜜网的攻击防护流程，按照欺骗场景生成部署、攻击诱捕、攻击行为分析、安全性增强等四种蜜网关键技术对现有研究成果进行分析归纳，详细讨论上述关键技术的作用及其研究进展，总结分析现有蜜网研究存在的问题与不足，展望未来的发展趋势和面临的挑战。

关键词　蜜网　欺骗防御　网络攻击　蜜罐

中图分类号　TP393　　**文献标志码**　A　　DOI:10.3969/j.issn.1000-386x.2024.10.053

DECEPTION DEFENSE ORIENTED HONEYNET TECHNIQUES

Liu Yaqun　Gao Yazhuo　Xing Changyou*　Zhang Guomin
(*Army Engineering University*, *Nanjing 210007*, *Jiangsu*, *China*)

Abstract　Honeynet deceives the attackers by constructing a trapping environment and masquerading as a real business network. Attracting attackers, monitoring attackers' behavior and analyzing their characteristics have become the trump card of network deception defense. The definition, classification and functions of honeynets were introduced. On this basis, combining the attack protection process of honeynets, the existing research results were analyzed and concluded according to the key technologies of honeynets, such as generation and deployment of deception scenarios, attack trapping, attack behavior analysis, and security enhancement. In addition, the effect and research progress of the above-mentioned key technologies were discussed in detail and the existing problems and shortcomings in the existing honeynet research were summarized. The development trend and challenge in the future were prospected.

Keywords　Honeynet　Deception defense　Cyber attack　Honeypot

0　引　言

近年来，网络攻击数量的不断增加以及网络攻击手段的不断丰富对网络环境的安全与稳定提出了严峻的挑战。据国家互联网应急中心抽样检测，2020 年上半年，我国境内约有 304 万台主机感染了计算机恶意程序，约 4 208 万个 IP 地址受到计算机恶意程序攻击，发现境内外约 1.8 万个 IP 地址对我国境内约 3.59 万个网站植入后门，国家信息安全漏洞共享平台（China National Vulnerability Database，CNVD）收录通用型安全漏洞 11 073 个[1]。网络攻击会造成大量隐私数据泄露，政府、企业的网站被篡改，大面积的网络瘫痪，给个人和国家造成难以估量的损失。

目前在网络防御中广泛使用的传统的网络防御手段（例如防火墙、入侵检测系统、入侵阻止系统等）通过匹配规则库来检测网络攻击，这些网络防御手段只能检测到已知的攻击，无法有效地发现并阻止高级可持续威胁（Advanced Persistent Threat，APT）等攻击。这一不足导致了 APT 攻击可以轻松地绕过传统的网络防御设备并入侵真实生产网络，网络资产的拥有者只有在蒙受巨大的损失后才能发现并制止 APT 等

收稿日期：2021－05－11。国家自然科学基金项目（61379149）；国家博士后科学基金项目（2017M610296）。**刘亚群**，硕士，主研领域：网络空间安全。**高雅卓**，硕士。**邢长友**，副教授。**张国敏**，副教授。

攻击。

欺骗防御技术作为一种主动式防御技术,通过主动对抗来达到防御效果,有希望扭转“易守难攻”的网络安全局势。欺骗防御技术并不尝试构建一个没有漏洞的系统,也不去刻意阻止具体的攻击行为,而是通过混淆、伪装等方法隐藏受保护系统的外部特征,使系统展现给攻击者的是一个有限甚至完全隐蔽或者错误的攻击面,导致攻击复杂度和攻击者代价增长。

蜜网作为欺骗性防御技术的一种,有效弥补了传统防御手段的不足,从一出现就引起了学术界和工业界的广泛关注。蜜网与欺骗防御的关系如图 1 所示。与基于规则库的传统的网络防御手段不同,蜜网可以在攻击者攻击成功之前就发现攻击者并阻止攻击者入侵生产网络。蜜网通过伪装成真实的生产网络来欺骗攻击者,吸引攻击者攻击并消耗攻击者的资源和时间,分析攻击者的攻击手段并了解攻击者的意图,从而达到保护真实生产网络的目的。从蜜网产生以来,蜜网已经在网络欺骗防御中应用并发展了约二十年时间。结果表明蜜网在检测攻击与提供攻击者信息方面已经取得了突出的成绩,包括大规模拒绝服务攻击的应对、蠕虫的检测、发现和跟踪僵尸网络、捕获和分析网络钓鱼等。从近几年蜜网研究成果的数量和质量来看,蜜网的发展已经进入到相对成熟的阶段。

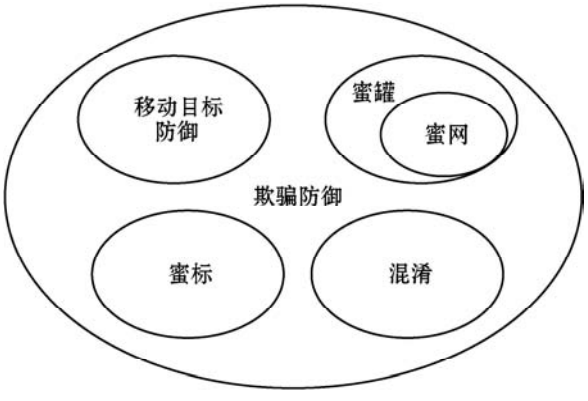

图 1　蜜网与欺骗防御的关系

虽然蜜网研究成果的数量与日俱增,但是目前专门介绍与总结蜜网的综述数量还相对较少。此外,现有蜜网综述[2]的研究与分析方法一般继承了蜜罐的讨论方法,这导致蜜网的功能和特点在很大程度上难以体现。为了弥补现有蜜网综述研究与分析方法的欠缺与不足,本文结合网络空间欺骗防御的思想梳理了近几年一些比较典型的蜜网研究成果,介绍了蜜网的概念与特点,归纳并整理了蜜网中常用的技术原理与实现手段,总结了当前蜜网研究工作的不足并展望了未来蜜网研究工作的发展趋势。

1　蜜网的相关概念

1.1　蜜网的定义

蜜网是由蜜罐演化而来,“蜜罐”的思想最早来源于《The Cuckoo's Egg》[3]一书,书中描述了一系列用于进行恶意行为分析的工具。随后,Lance Spitzner 首先给出了蜜罐的定义:蜜罐是一种安全资源,其价值体现在被探测、攻击或者攻陷的时候[4]。蜜罐通过伪装成真实的服务系统来欺骗攻击者,吸引攻击者攻击并分析攻击者的行为,从而达到保护真实网络的目的。虽然蜜罐已经具备欺骗防御功能,但是单一蜜罐的功能是有限的,难以满足实际需求。例如,低交互蜜罐虽然资源消耗少,方便部署,但是只模拟了简单的服务,可信度低,容易被攻击者识别;高交互蜜罐虽然模拟了完整的操作系统,可信度高,但是资源消耗大,无法大规模部署。

为了解决单一蜜罐功能受限的问题,研究人员开始将不同功能的蜜罐进行组合,形成蜜网。蜜网是对蜜罐的拓展与延伸,The Honeynet Project 将蜜网定义为:蜜网是由若干个能收集和交换信息的蜜罐构成的一个网络体系架构[5]。这一定义着重关注了蜜罐在蜜网中的作用。实际上,随着蜜网的不断发展,蜜网已经不再局限于蜜罐的简单组合,它还包含了入侵检测系统、网关等其他类型的网络防御设备和数据处理工具。基于以上描述,我们将蜜网定义为:蜜网是由不同种类的蜜罐、网关、入侵检测系统等网络防御设备、数据收集设备和数据分析设备按照一定的拓扑构成的网络体系架构。蜜网集成了不同网络防御设备和数据处理设备的优势,可以改变网络攻防博弈中攻强守弱的局面,并对维护安全的网络环境发挥重要作用。

1.2　蜜网的分类

到目前为止,已经有很多研究[6-7]提出了蜜罐的分类方法。由于蜜罐是蜜网最重要的组成部分,所以现有的蜜网分类方法基本都继承了蜜罐的分类方法。本节将现有的蜜网分类方法进行了总结,并列举了一些典型的蜜网系统,如表 1 所示。

表 1　传统的蜜网分类方法

分类依据	分类结果	代表蜜网
资源类型	物理蜜网	文献[8]
	虚拟蜜网	文献[9,11-12]
	混合蜜网	文献[13]

续表 1

分类依据	分类结果	代表蜜网
部署方式	静态蜜网	文献[8,14]
	动态蜜网	文献[9,15]
目的用途	生产蜜网	文献[8,16]
	研究蜜网	文献[17]
交互程度	低交互蜜网	文献[9]
	高交互蜜网	文献[8,11]
	混合蜜网	文献[18,19]

根据蜜网使用的资源类型蜜网可以分为物理蜜网、虚拟蜜网和混合蜜网。物理蜜网中的蜜罐是网络中真实的物理主机,例如 Gen I、Gen II[8] 等传统蜜网。物理蜜网可以给攻击者提供一个与真实服务系统完全相同的诱捕环境,可信度高,但是资源消耗多,难以大规模部署,这一缺点也导致物理蜜网的发展基本停滞。虚拟蜜网则使用虚拟化技术来部署蜜罐,目前常用的虚拟化技术有虚拟化软件、虚拟机和虚拟化容器等。Wang 等[9] 使用 honeyd[10] 软件来部署虚拟蜜网,Potemkin[11] 使用 Xen 虚拟机来部署虚拟蜜网,Honey-System[12] 则使用 docker 容器来部署虚拟蜜网。相对于物理蜜网,虚拟蜜网资源消耗少,易于部署和维护,但虚拟化环境容易被攻击者识别,可信度低。为了解决物理蜜网和虚拟蜜网各自存在的缺点,Artail 等[13] 将 honeyd[10] 虚拟蜜罐和真实物理主机相结合形成了混合蜜网,其中 honeyd 模仿生产网络中的操作系统和服务,吸引攻击者攻击,真实物理主机运行真实的服务与攻击者进行交互。

根据蜜网的配置方式蜜网可以分为静态蜜网和动态蜜网。静态蜜网需要安全人员在蜜网运行之前确定蜜网的配置,一旦蜜网运行,安全人员需要手动配置蜜网。静态蜜网对所有外部入侵事件采取相同的应对策略,隐蔽性差并且手动配置耗时耗力。例如 Gen I、Gen II、Collapsar[14] 等传统蜜网都必须在运行之前就完成蜜网的配置。动态蜜网能够根据外部攻击者的攻击方式和环境的变化实时更新配置,对外部入侵事件做出不同的响应。为了吸引更多的攻击者进入蜜网,Wang 等[9] 利用进入蜜罐的 ICMP 数据包的数量来定量描述蜜罐对于攻击者的吸引力,随后提出合作学习算法、进化算法和混合算法,这三种算法根据不同类型蜜罐对攻击者的吸引力来动态更新蜜网中蜜罐的类型。为了减少资源消耗,Elastic-hybrid[15] 能够在攻击者进出蜜网时动态地生成和销毁高交互蜜罐,实现蜜网的动态配置功能。

按照蜜网的目的蜜网可以分为生产蜜网和研究蜜网。生产蜜网通常部署在公司或者企业的生产网络旁边,目的是吸引攻击者,减少公司或企业的生产网络被攻击的概率,增强公司或企业生产网络的安全性。例如 Gen I、Gen II、Dressed up[16] 等蜜网都部署在生产网络周围用于吸引攻击者攻击,这样可以减少生产网络被攻击的概率。研究蜜网通常由研究人员或组织开发、利用,目的是收集攻击者的信息,分析攻击者的动机、行为、使用的工具等,并寻找应对这些攻击的防御方法。Honeycomb[17] 对 honeyd[10] 进行了拓展,通过报头对比和有效负载分析产生攻击者的攻击特征,是一个典型的研究蜜网。

根据蜜网的交互程度把蜜网可以分为低交互蜜网、高交互蜜网和混合蜜网。低交互蜜网使用低交互蜜罐来部署蜜网,Wang 等[9] 使用 honeyd[10] 低交互蜜罐来部署蜜网,此外,Dionaea、Glastopf、Kippo 等都是常见的低交互蜜罐。低交互蜜网资源消耗少,易于大规模部署,但是与攻击者的交互能力有限,可信度低,容易被攻击者识别。高交互蜜网使用虚拟机、物理主机等高交互蜜罐来部署蜜网,例如 Gen I、Gen II 使用真实的物理主机来部署蜜网,Potemkin[11] 使用 Xen 虚拟机部署蜜网。高交互蜜罐逼真度高,但资源消耗多,不便于大规模部署。为有效平衡蜜网的逼真度和部署成本,混合蜜网同时使用低交互蜜罐和高交互蜜罐来部署蜜网,混合蜜网中的低交互蜜罐一般用于扩大攻击面,大规模地吸引攻击者,高交互蜜罐一般用于给攻击者提供一个高逼真的诱捕环境,对有研究价值的攻击者进行深入分析。VMI-Honeymon[18] 同使用 Dionaea 低交互蜜罐和 Xen 高交互蜜罐来部署混合蜜丸,Fan 等[19] 则同时使用 honeyd 低交互蜜罐和 VNX 高交互蜜罐来部署混合蜜网。

除了以上四种常见的蜜网分类方式之外,Fan 等[2] 提出了一些比较特殊的蜜网分类方法。根据蜜网中蜜罐 IP 地址的范围,蜜网可以分为独立蜜网和分布蜜网,独立蜜网中蜜罐的 IP 地址都属于同一网段,分布蜜网中蜜罐的 IP 地址属于不同网段。根据蜜网的部署范围,蜜网可以分为本地蜜网和远程蜜网,本地蜜网中的蜜罐都部署在同一区域内,远程蜜网中的蜜罐部署在不同区域。根据蜜网与生产网络的逻辑关系,蜜网可以分为雷区蜜网和保护蜜网,雷区蜜网中的蜜罐部署在生产网络当中,而在保护蜜网中,蜜网是生产网络的镜像。

1.3 蜜网的主要功能

根据实际需求和用途的不同,各蜜网系统通常具备不同的功能。通过对现有的蜜网进行分析与总结,

我们将蜜网的功能概括为以下四种:攻击欺骗、数据控制、数据捕获和攻击分析。

攻击欺骗是指欺骗攻击者,使攻击者认为蜜网是一个真实的生产网络,从而对蜜网发起攻击。蜜网主要通过模拟真实的操作系统和服务来欺骗并吸引攻击者,因此模拟的真实性是影响蜜网欺骗效果的重要因素。

数据控制是指对进入蜜网的流量进行适当的处理并采取一系列措施,防止攻击者在取得蜜网的控制权限后将蜜网作为一个僵尸网络来攻击真实网络中的生产系统。数据控制功能通常由蜜墙、SDN 交换机等网关设备来实现,数据控制可以提高攻击信息的捕获效率并保证蜜网系统的安全性。

数据捕获是在攻击者进入蜜网后,蜜网通过与攻击者进行交互来记录攻击者的攻击行为,收集攻击者的信息。数据捕获功能可以通过流量监控、文件监控、内存监控、进程监控等多种方式实现。此外,蜜网需要保证数据捕获行为和捕获的攻击数据的隐蔽性,防止攻击者通过数据捕获的行为和结果来识别蜜网。

攻击分析是对收集的攻击者信息进行分析,提取攻击者的攻击特征,并提出相应的攻击防御方法。攻击分析通常需要借助专门的数据分析技术和工具,近年来机器学习技术在攻击分析领域得到了广泛的应用并取得了不错的效果。

2　蜜网关键技术分析

除蜜罐之外,当前的蜜网还包含其他网络防御设备和数据处理设备,因此蜜网具备比蜜罐更加丰富的功能和特点。为了对蜜网的功能和特点进行更加细致深入的研究与分析,本部分结合蜜网在网络攻击防护流程中不同阶段的功能,结合蜜网的欺骗场景生成部署、攻击诱捕、攻击行为分析、安全性增强四种关键技术对现有成果进行分析讨论。

2.1　欺骗场景生成部署技术

欺骗场景生成部署技术的主要任务是部署诱捕环境,为实现精准高效的攻击诱捕奠定基础。部署诱捕环境可以欺骗攻击者,吸引攻击者攻击,进而减小业务系统被攻击的概率。

欺骗场景生成部署技术主要分为静态欺骗场景生成部署技术和动态欺骗场景生成部署技术。早期蜜网大多采用静态欺骗场景生成部署的方式,这一方式需要在蜜网运行前手工完成蜜网的搭建与设置,且蜜网在运行过程中无法动态地调整蜜网的配置。由于攻击者通常不会重复地选择并入侵同一目标,因此基于静态部署的蜜网对于攻击者的吸引力会随时间的流逝而逐渐下降。为解决这一弊端,一些蜜网逐渐开始采用动态欺骗场景生成部署的方式,在蜜网运行过程中动态地调整蜜网的配置。动态部署的方式大致可以分为两种,一种是根据受保护网络中各类型节点的数量动态调整诱捕环境中各类型节点的数量,其工作流程如图 2 所示。Hecker 等[20]将基于 p0f 和 tcpdump 的被动扫描和基于 nmap 和 xprobe2 的主动扫描相结合,提出了被动、低等、中等、中高等和高等五种不同的扫描方式以及完全复制受保护网络、部分复制受保护网络和与受保护网络结合三种不同的蜜网配置方法,将扫描结果存放在数据库中并根据扫描结果动态配置蜜网。Fan 等[21]提出的 Honeyvers 支持使用 nmap 工具扫描受保护网络并将扫描结果保存为 XML 文件,随后使用其自行设计的 TIHDL 语言格式化扫描数据并生成蜜网的配置。近年来,机器学习在蜜网部署领域的运用也取得了不错的效果,Fraunholz 等[22]使用 nmap 工具扫描受保护网络获取受保护网络的相关信息,随后根据其中的 TCP 指纹和开放端口号等信息,使用动态的 k-均值聚类算法对网络聚类并生成蜜网配置,动态部署蜜网。蜜网动态部署的另外一种方式是根据诱捕环境中不同类型节点对于攻击者的吸引力大小来动态调整诱捕环境中各类型节点的数量。Wang 等[9]根据进出蜜网的 ICMP 数据包的数量来确定蜜罐对攻击者的吸引力,并设计三种蜜网部署算法:合作学习算法、进化算法和混合学习算法,动态更新蜜网中不同类型蜜罐的数量。王航[23]提出一种基于反馈的蜜网动态配置算法,蜜网感知网络环境变化并根据网络攻击成功与否给出反馈信号,系统依据反馈信号更新蜜网的配置策略。

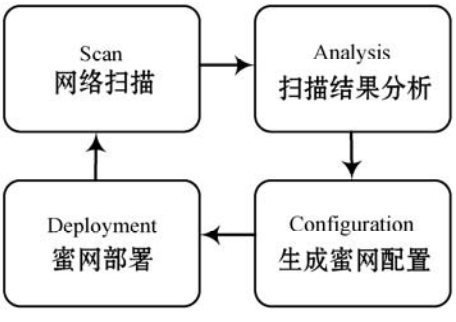

图 2　蜜网动态配置工作流程

2.2　攻击诱捕技术

攻击诱捕技术的主要任务是吸引攻击者并对进入蜜网的流量进行预处理,以欺骗的方式干扰攻击者的认知和行为,使攻击者误以为自己成功入侵真实系统,收集攻击者的相关数据并对攻击特征进行分析。攻击

诱捕技术主要包括攻击分类、拓扑混淆、攻击迁移和攻击监控。

2.2.1 攻击分类

攻击分类利用传统的防火墙、入侵检测系统、数据包分析和蓬勃发展的人工智能等技术,识别恶意流量并分析恶意流量的研究价值,对外部流量和其中的恶意流量进行分类。外部流量根据其是否存在异常可分为正常流量和异常流量。未发现异常的流量通常由真实系统提供相应服务,发现异常的流量则由诱捕节点响应并实施诱捕。为保护网站正常运行,林建宝[24]使用基于规则匹配的流量异常检测和基于威胁情报库检测两种方法检测外部流量,将检测出的异常流量和可疑流量转发至影子服务,并将正常流量转发至正常网站。对于外部流量中的恶意流量,根据其研究价值可分为两类,低价值流量和高价值流量。研究价值较低的恶意流量由低交互蜜罐响应或直接丢弃,研究价值高的恶意流量则由高交互蜜罐响应。Fan 等[19]使用 honeybrid 网关将攻击者分为匹配特定指纹的攻击(已知攻击)、包含未知内容的攻击(零日攻击)、低交互蜜罐无法处理的攻击,并将具有研究价值的攻击流量迁移至高交互蜜罐中,snort 作为 honeybrid 的拓展可以决策是否丢弃研究价值较低的攻击流量。由于 snort 的检测规则编写灵活,网络安全人员可以结合实际需求自行设计攻击检测规则。近年来有蜜网使用人工智能技术进行攻击检测与分类,取得了不错的效果,在将来具有一定的发展潜力。针对 SSH 攻击,Zuzcak 等[25]自定义攻击者的评估特征并使用 EMYCIN 专家系统来评估攻击者的研究价值,研究价值较低的攻击者由中交互蜜罐响应,研究价值较高攻击者则迁移至高交互蜜罐响应,以有限的资源实现有效的攻击分类。网络扫描流量是攻击者在侦察阶段发送的流量,用于攻击者获取网络情报,这部分流量特征明显,因此属于研究价值较低的恶意流量。一些蜜网将识别到的网络扫描流量发往到由低交互蜜罐组成的拓扑混淆模块处理,以极低的成本欺骗攻击者,干扰攻击者的认知,吸引攻击者攻击,如图 3 所示。王贺[26]和李俨[27]均利用攻击检测模块检测低级别大范围的扫描探测流量并使用软件定义网络技术将检测出的扫描探测流量转发至拓扑模拟蜜罐,进行攻击吸引。除了根据研究价值对恶意流量进行分类外,一些蜜网还根据恶意流量的目的端口号进行分类,将具有不同目的端口号的流量转发至提供相应服务的蜜罐上,如图 4 所示。Stoecklin 等[16]使用欺骗路由器检查数据包的目的端口号,如果攻击者试图访问真实网络系统中不存在的服务,欺骗路由器会将其发出的流量转发至蜜罐中。HoneyMix[28]在其蜜网的每个蜜罐上部署多种服务,其服务映射表记录了每个蜜罐提供的服务种类,其转发决策引擎负责根据服务映射表将攻击流量转发到所有提供相关服务的蜜罐中,提高资源的利用率。

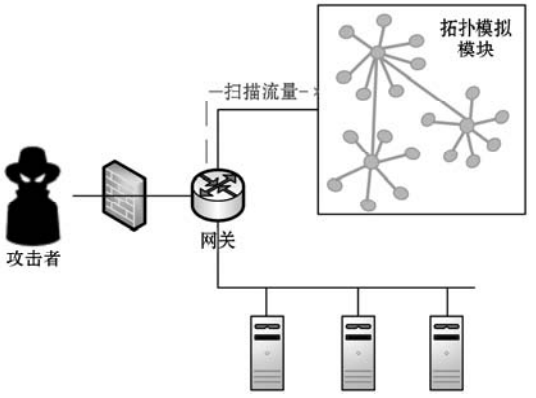

图 3 基于扫描流量的攻击分类

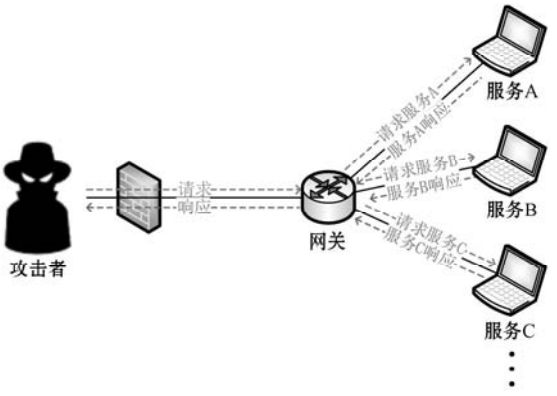

图 4 基于目的端口号的攻击分类

2.2.2 拓扑混淆

文献[29]指出 70% 的攻击活动前都存在侦察行为,攻击者在攻击的早期利用侦察可以获取目标网络的信息,进而选择攻击的目标和手段。拓扑混淆以欺骗的方式给攻击者提供虚假的网络信息,隐藏真实网络并吸引攻击者进入诱捕环境。拓扑混淆的实现方式分为两种,一种是利用虚拟化技术创建多个虚拟节点来模拟真实的网终端点,形成虚拟网络。Shing 等[30]将恶意流量重定向到模拟网络上真实终端行为的网络 tarpits 上,创建粘性连接来减缓或阻止自动扫描和迷惑对手。为了应对 LDoS 攻击,SDHoneyNet[31]找到受保护网络的瓶颈链路并使用 Minine 在相关节点上创建复杂的虚拟拓扑,通过将 traceroute 数据包重定向至虚假网络使攻击者无法找到瓶颈链路。另一种拓扑混淆的实现方式是伪造数据包回复攻击者发送的扫描流量或修改真实数据包的相关字段来误导攻击者。在 Achleitner 等[32]设计的侦察欺骗系统中,欺骗服务器负责根据欺骗策略伪造网络数据包回复攻击者,给不同位置的攻击者提供不同的虚假拓扑信息,可以有效延长脆弱主机被发现的时间。Wang 等[33]利用软件定

义网络技术灵活的可编程性解析攻击者发送的扫描流量并根据指纹库和虚假拓扑信息伪造相应数据包来回复攻击者,给攻击者提供虚假的拓扑信息,起到大范围吸引攻击者的作用。Zhang 等[34]利用“球桶模型”计算虚拟节点对于真实系统的保护效果,通过 SDN 交换机伪造数据包来创建虚拟网络拓扑并动态变换其中虚拟节点的类型,在混淆真实系统的同时可以有效地发现入侵的攻击者。

2.2.3　攻击迁移

攻击迁移能以对攻击者透明的方式改变攻击诱捕的位置,对攻击者实施精准诱捕,在减少资源消耗的同时可以增大诱捕成功率。

攻击迁移包含攻击流量的迁移和攻击状态的迁移,攻击流量的迁移通常利用网络代理、软件定义网络等技术在网络层改变攻击流量的转发方向,攻击状态的迁移则通常利用流量重放、虚拟化迁移等技术迁移攻击者在服务端的状态,其中攻击状态的迁移是决定攻击迁移过程对攻击者是否透明的关键。攻击迁移的场景可以分为两种,第一种是从低交互蜜罐到高交互蜜罐的攻击迁移,如图 5 所示。针对 SSH 攻击,Sun 等[35]使用前端诱饵拦截攻击者发送的所有恶意指令,随后利用 SSH 隧道将拦截的恶意指令通过网关发送至后端诱饵,最后前端诱饵接受后端诱饵的响应信息并加上新的数据包头部,回复攻击者,兼顾蜜网的规模性和可扩展性。HoneyDOC[36]提出一种基于 TCP 三次握手重放的攻击状态迁移方法,它利用修改过源代码的 ofsoftswitch13 交换机修改 TCP 数据包头部的 seq 和 ack 值,在攻击者与高交互蜜罐之间重放攻击者与低交互蜜罐之间的 TCP 三次握手过程,保证攻击者的攻击状态不中断地从低交互蜜罐迁移至高交互蜜罐。另外,HoneyDOC 利用软件定义网络技术实现攻击流量迁移的功能。不足的是 HoneyDOC 必须在攻击者与低交互蜜罐之间完成 TCP 三次握手后立即判断数据包的利用价值并决定是否实施攻击迁移。如果攻击者与低交互蜜罐之间进行了长时间交互并且改变了低交互蜜罐的状态,蜜网系统必须记录攻击者发送给低交互蜜罐的所有数据包并在攻击者与高交互蜜罐之间重放这些数据包[37]。攻击迁移的第二种场景是从真实服务器到诱捕节点的攻击迁移,如图 6 所示。INTERCEPT +[38]修改 QEMU 的源代码,基于 QEMU 的虚拟机实时迁移技术完成对攻击者攻击状态的透明迁移,同时保证真实系统在迁移之后能够继续运行,利用软件定义网络技术实现攻击流量的迁移功能。Honey-Patches[39]拓展 CRIU 的源码实现 LXC 容器的实时迁移功能,将试图攻击真实系统上诱饵漏洞的攻击者迁移到诱捕环境中,利用反向代理实现攻击流量的迁移功能,其资源消耗少,适合大规模部署。Sandnet[40]则利用 CRIU 技术实现 Docker 容器的实时迁移功能,并且将所有与被入侵节点建立连接的真实节点都迁移到诱捕环境中,利用软件定义网络技术迁移所有入侵真实节点的攻击流量。相比于虚拟机迁移,容器迁移的速度更快,因此具有更高的隐蔽性,具有较好的应用前景。

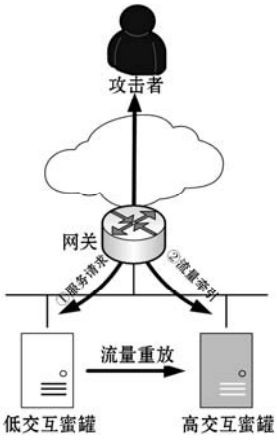

图 5　低交互蜜罐到高交互蜜罐的攻击迁移

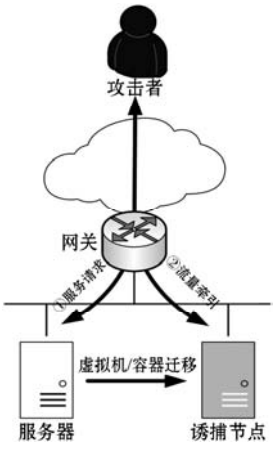

图 6　服务器到诱捕节点的攻击迁移

2.2.4　攻击监控

攻击监控的主要任务是收集攻击流量或监控攻击者在真实系统或诱捕环境中的行为,为后续提取攻击者的攻击特征做准备。流量收集是最简单的攻击监控方式,Wang 等[41]使用 tcpdump 配置参数并收集进入蜜罐的流量,tcpdump 可以将收集的数据输出到控制台或存储为 pcap 等格式的文件。Chuang 等[42]利用安装在 OpenFlow 交换机内的蜜罐收集外部流量并保存到文件中。除了流量收集之外,早期的蜜网通常使用 Sebek 等特定的攻击监控软件来实现攻击行为的监控功能。但是由于监控软件安装在蜜罐内部,在安装监控软件的蜜罐被攻击者攻陷之后,这些监控软件很容易被攻击者发现。为了解决这一问题,虚拟机自省技术开始逐渐运用到蜜网的攻击监控中。虚拟机自省在

虚拟机监视器(Virtual Machine Monitor,VMM)监控虚拟机内部的系统调用事件,它可以提供与 Sebek 同等程度的监控能力,并且不容易被攻击者发现。Lengyel 等[18]基于多个开源的自省和取证工具设计并实现 VMI-Honeymon,VMI-Honeymon 对被监控的虚拟机透明并且利用内存扫描克服对虚拟机内核的依赖,能有效监控高交互蜜罐,捕获已知和未知的恶意软件。Urias 等[43]使用专为 KVM 虚拟机管理程序设计的 KVMi 技术,KVMi 作为一个可加载的内核模块,允许攻击者完全控制虚拟机,并且在不引入明显改动的情况下监控攻击者在虚拟机内部的行为,保证攻击监控的隐蔽性。与虚拟机自省不同,蜜标是一种欺骗型的攻击行为监控方式。蜜标是一种用来吸引攻击者获取并使用的信息资源,正常情况下的合法用户并不会试图利用蜜标,部署在真实系统中的蜜标具有发现攻击者的作用。蜜标可以误导攻击者,使攻击朝向对防御者有利的方向发展。利用蜜标,网络管理人员可以轻松获取攻击者的位置,监控攻击者的行为。Araujo 等[39]在真实系统上部署诱饵漏洞,所有试图攻击这一漏洞的攻击者都会暴露自己的身份,进而被迁移到诱捕环境中。为保护电网,DefRec[44]向其中的诱饵节点发送诱饵流量,增加攻击者被动网络扫描侦察的难度,伪造诱饵数据作为虚拟节点发送的网络数据包的有效负载,误导攻击者实施无效的攻击。

2.3 攻击行为分析技术

攻击分析技术主要负责分析蜜网捕获的攻击流量和攻击行为,提取攻击者的攻击特征,为真实网络系统提供攻击防御方法。

McGrew 等[45]认为攻击者的攻击特征包含动机、广度/深度、复杂度、隐蔽性、攻击源、漏洞、工具等信息,各信息的含义如表 2 所示。在此基础上,Nawrocki 等[6]介绍了攻击来源、攻击目标、攻击频率、攻击演化、攻击繁殖、攻击模式、攻击根源、攻击风险评估、漏洞检测等攻击信息。其中攻击来源、攻击目标、攻击频率可以从攻击日志中直接获得,因此大多数研究者都会分析这三个攻击特征。其余的攻击信息需要通过复杂的分析来获得,因此相关的分析方法出现相对较晚。

表 2 攻击特征及含义

攻击特征	含义
动机	攻击者实施攻击的原因
广度/深度	广度表示被攻击主机的数量,深度表示攻击者对特定主机造成的损害
复杂度	攻击者实施攻击需要具备的专业知识
隐蔽性	攻击者销毁攻击证据的能力
攻击源	识别攻击者或攻击的来源
脆弱点	攻击者实施攻击所利用的漏洞
工具	攻击者实施攻击所使用的工具

Al-Mohannadi 等[46]定义的威胁情报包含三部分:Attack、Behaviour 和 Pattern,使用 Elastic Stack 实现蜜罐数据分析和可视化,产生威胁情报,可以帮助防火墙、IDS 和 IPS 等保护生产网络。Elastic Stack 是一个免费开源的数据收集与分析工具,Elastic Stack 由 Elasticsearch、Kibana、Beats 和 Logstash 组成,能够安全可靠地获取数据并实时地对数据进行搜索、分析和可视化。Elastic Stack 可以帮助蜜网快速有效地收集并分析攻击数据,获得理想的攻击信息和攻击特征。

在早年,规则挖掘、内存污点等技术在蜜网的攻击分析中得到了广泛应用。近年来,越来越多的研究者将机器学习应用于蜜网的攻击分析阶段,识别攻击流量,提取攻击特征,取得不错的效果。El Kamel[47]使用 K-means 算法对进入蜜罐的流量进行聚类,然后使用线性回归算法对每一个类别进行建模,进而将蜜罐中的流量分为正常流量和恶意流量。Boukela 等[48]使用离群点检测算法 Local Outlier Factor(LOF)对蜜罐中的流量进行异常检测并利用 outlier ensem-bles 将具有不同参数和不同数据子空间的 LOF 的执行结果进行集成,解决了数据的维度灾难和子空间中的异常隐藏等问题。

2.4 蜜网安全性增强技术

由于蜜网需要主动吸引攻击者攻击,因此蜜网的真实性成为影响蜜网诱捕效果的关键因素。除此之外,蜜网不对外提供正常服务,所以蜜网不具备任何生产价值,不需要防御外来入侵。但是,大多数蜜网是由真实系统组成,入侵蜜网的攻击者一旦获得蜜网的控制权,往往会把蜜网当作僵尸网络来攻击其他真实的生产系统,对生产网络构成巨大威胁。蜜网安全性增强技术的主要任务正是提高蜜网的真实性,并且应对攻击者在蜜网内部发起的横向渗透。

蜜网的逼真度是决定诱捕是否成功的关键,数据包的响应时间、服务的完整性、蜜罐的指纹等都可以被攻击者用来进行蜜网的识别[49-50]。通常情况下,虚拟化蜜罐的回复时延高于真实系统,OpenFire[51]指出在大多数情况下,只需要给真实的物理系统增加小于 1 毫秒的时延,就可以极大地提高蜜网的隐蔽性。使用

专门的硬件设备运行蜜罐可以减小蜜网的回复时延,同样可以提高蜜网的隐蔽性。攻击者利用数据包的相关字段可以区分蜜罐与真实系统,传统的指纹隐藏设备可隐藏蜜罐的指纹信息,减小蜜网被攻击者识别的概率。HoneyMix[28]的回复清洗模块基于已知的指纹技术修改回复报文中可能被攻击者利用的字段,减少了蜜网系统被识别的可能性。类似地,HoneyProxy[52]的回复处理器负责检查蜜罐回复的报文是否含有蜜网的指纹信息,其连接管理引擎负责从队列中选择最安全的报文来回复攻击者。Naik 等[53-54]分别使用主成分分析、动态模糊规则插值等方法有效地识别针对蜜罐的指纹识别数据包,保证了蜜罐的逼真性与隐蔽性。近年来,强化学习在蜜网中的应用有效提高了蜜网的隐蔽性。Hayatle[55]将蜜罐对攻击的处理过程建模成一个部分可观测马尔可夫决策过程,其蜜罐有 waiting、compromised 和 disclosed 三种状态以及 allow、not-allow 和 reset 三种动作,它能根据外部攻击者和环境的变化采取不同的响应策略,提高蜜网的隐蔽性并延长攻击者与蜜网的交互时间,收集更多的攻击者信息。Dowling 等[56]规定了 allow、block 和 substitute 三种蜜罐动作并使用 SARSA 强化学习算法来寻找蜜罐对于僵尸攻击的最佳响应策略,延长了蜜罐与僵尸攻击的交互时间。

为了防止蜜网成为攻击者实施攻击的跳板,攻击限制可以延缓攻击者的攻击速率或者改变出站流量的方向,保护真实的网络资产。Gen Ⅰ的 Honeywall 限制了蜜网出站流量所建立的连接的数量,Gen Ⅱ的 Honeywall 则修改了蜜网出站报文的内容使得攻击失效。Collapsar 的限制模块通过限制出站数据包的发送速率来限制蜜网的出站流量。流量限速虽然可以限制攻击者将蜜罐作为跳板来攻击真实服务器,但会导致蜜网与正常生产网络的行为模式产生差异,从而引起攻击者的怀疑。攻击者察觉到这一差异后会减少甚至停止自己的攻击行为,导致蜜网无法完成既定的目标。流量重定向可以将蜜网的出站流量重定向到蜜网新的蜜罐中,使攻击者误以为进入新的脆弱系统,从而对攻击者实施进一步诱捕。Potemkin[11]的网关负责对出站流量进行检测,如果出站流量的目的地址超出限定的范围,Potemkin 会将这些流量重定向到蜜网中新的蜜罐上。与流量限速相比,流量重定向具有明显的优势。它不仅能保证蜜网的隐蔽性,而且能在蜜网中捕获蠕虫传播的整个过程,不影响蜜网的正常运行。

3　研究展望

随着主动防御、欺骗防御等理念的发展,蜜网作为当下一种热门的欺骗防御技术得到了越来越多的关注,围绕构建更加逼真的攻击欺骗场景,以及更加智能的攻击诱捕模型等,蜜网在体系架构、资源利用、技术运用等方面仍将得到持续发展。

在蜜网架构与资源利用方面,结合了低交互蜜罐可扩展性以及高交互蜜罐高逼真性的混合蜜网实现了规模与性能的平衡,仍将是未来蜜网的主要发展趋势。然而,攻击者的状态在从低交互蜜罐到高交互蜜罐的迁移前后的一致性问题在混合蜜网中仍然没有得到彻底解决,混合蜜网的真实性和隐蔽性难以得到保障。有必要结合 SDN、轻量级虚拟化等技术的发展,进一步研究提高欺骗诱捕效能的蜜网架构。

在技术运用方面,一些热门的技术在蜜网中有比较好的应用前景。

(1) 蜜网与软件定义网络技术相结合。软件定义网络技术将传统路由设备的数据层和控制层解耦,蜜网管理人员可以通过编写控制器程序的方式来灵活控制蜜网中的流量转发行为。此外,基于软件定义网络技术的蜜网可以通过修改报文的相关字段来应对攻击者的指纹识别技术,增强蜜网的隐蔽性。

(2) 蜜网与机器学习相结合。机器学习在蜜网中有很多应用场景,在攻击检测时,神经网络、聚类分析等方法不需要构建规则库并能提高检测的准确率。在动态响应时,强化学习可以帮助蜜网不断改进配置以提高蜜网的吸引力。

(3) 蜜网与虚拟化技术相结合。虚拟蜜网部署成本低,易于维护和管理,并能有效地监控攻击者的攻击行为,虚拟化技术在蜜网的发展过程中发挥了重大作用。然而,CPU 时钟采样、指令执行时间对比、命名空间检测等环境检测技术可以轻易地识别虚拟环境,网络攻击者在虚拟环境中会变得十分谨慎,这在一定程度上降低了虚拟蜜网的欺骗效果。因此,增强虚拟蜜网的隐蔽性是一个重要的研究方向。

4　结　语

作为一种欺骗防御技术,蜜网弥补了传统网络防御技术的不足,是目前网络安全研究的热点领域。本文首先从定义、分类和功能的角度介绍了蜜网的相关概念,随后对于蜜网的关键技术,本文在欺骗场景生成部署、攻击诱捕、攻击行为分析、安全性增强四个方面对蜜网的系统架构、功能特点、实现原理、技术手段等进行了分析与归纳,总结了蜜网的优势与不足,并对蜜网下一步的研究趋势进行了展望。

除了传统计算机网络之外,蜜网在物联网、工业网

络等其他领域也有广泛的应用前景。随着新技术、新领域与蜜网的有效结合,蜜网将得到不断发展更新,为构建更加安全的网络环境发挥更重要的作用。

参 考 文 献

[1] 国家互联网应急中心. 2020 年上半年我国互联网网络安全监测数据分析报告[EB/OL]. [2021 - 05 - 11]. https://www.cert.org.cn/publish/main/46/2020/20200926085042652505447/20200926085042652505447_.html.

[2] Fan W J, Du Z H, Fernandez D. Taxonomy of honeynet solutions[C]//SAI Intelligent Systems Conference, 2015: 1002 - 1009.

[3] Stoll C. The cuckoo's egg: Tracking a spy through the maze of computer espionage[M]. New York: Simon and Schuster, 2005.

[4] Spitzner L. Honeypots: Catching the insider threat[C]// 19th Annual Computer Security Applications Conference, 2004: 170 - 179.

[5] The honeynet project. Know your enemy: Learning about security threats[M]. 2nd ed. Boston: Addison-Wesley Professional, 2004.

[6] Nawrocki M, Wählisch M, Schmidt T C, et al. A survey on honeypot software and data analysis[EB]. arXiv: 1608.06249, 2016.

[7] Fan W J, Du Z H, Fernandez D, et al. Enabling an anatomic view to investigate honeypot systems: A survey[J]. IEEE Systems Journal, 2018, 12(4): 3906 - 3919.

[8] Spitzner L. The honeynet project: Trapping the hackers[J]. IEEE Security & Privacy, 2003, 1(2): 15 - 23.

[9] Wang H F, Chen Q K. Dynamic deploying distributed low-interaction honeynet[J]. Journal of Computers, 2012, 7(3): 692 - 698.

[10] Provos N. A virtual honeypot framework[C]//13th Conference on USENIX Security Symposium, 2004: 13.

[11] Vrable M, Ma J, Chen J, et al. Scalability, fidelity, and containment in the Potemkin virtual honeyfarm[C]//20th ACM Symposium on Operating Systems Principles, 2005: 148 - 162.

[12] Majithia N. Honey-system: Design implementation attack analysis[D]. Kanpur: Indian Institute of Technology Kanpur, 2017.

[13] Artail H, Safa H, Sraj M, et al. A hybrid honeypot framework for improving intrusion detection systems in protecting organizational networks[J]. Computers & Security, 2006, 25(4): 274 - 288.

[14] Jiang X, Xu D. Collapsar: A VM-based architecture for network attack detention center[C]//13th Conference on USENIX Security Symposium, 2004: 1907.

[15] Bao N K, Ahn S, Park M. An elastic-hybrid honeynet for cloud environment[J]. Computer Science & Information Technology, 2018, 42(1): 117 - 127.

[16] Stoecklin M P, Zhang J L, Araujo F, et al. Dressed up: Baiting attackers through endpoint service projection[C]// ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization, 2018: 23 - 28.

[17] Kreibich C, Crowcroft J. Honeycomb: Creating intrusion detection signatures using honeypots[J]. ACM SIGCOMM Computer Communication Review, 2004, 34(1): 51 - 56.

[18] Lengyel T K, Neumann J, Maresca S, et al. Virtual machine introspection in a hybrid honeypot architecture[C]// 5th Workshop on Cyber Security Experimentation and Test, 2012.

[19] Fan W J, Du Z H, Fernandez D. Adaptive and flexible virtual honeynet[C]//1st International Conference on Mobile, Secure and Programmable Networking, 2015: 1 - 17.

[20] Hecker C, Hay B. Automated honeynet deployment for dynamic network environment[C]//46th Hawaii International Conference on System Sciences, 2013: 4880 - 4889.

[21] Fan W J, Fernandez D, Du Z H. Versatile virtual honeynet management framework[J]. IET Information Security, 2017, 11(1): 38 - 45.

[22] Fraunholz D, Zimmermann M, Schotten H D. An adaptive honeypot configuration, deployment and maintenance strategy[C]//19th International Conference on Advanced Communication Technology, 2017: 53 - 57.

[23] 王航. 动态蜜网关键技术研究与实现[D]. 成都:电子科技大学,2019.

[24] 林建宝. 基于网络欺骗的网站防护技术研究[D]. 北京:北京邮电大学,2018.

[25] Zuzcak M, Zenka M. Expert system assessing threat level of attacks on a hybrid SSH honeynet[J]. Computers & Security, 2020, 92: 101784.

[26] 王贺. 基于 SDN 的混合蜜网系统设计与实现[D]. 北京:北京邮电大学,2019.

[27] 李俨. 基于 SDN 的蜜网主动防御系统设计与实现[D]. 北京:北京邮电大学,2019.

[28] Han W, Zhao Z M, Doupe A, et al. HoneyMix: Toward SDN-based intelligent honeynet[C]//ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization, 2016: 1 - 6.

[29] Chakravarty S, Portokalidis G, Polychronakis M, et al. Detecting traffic snooping in tor using decoys[J]. International Workshop on Recent Advances in Intrusion Detection, 2011: 222 - 241.

[30] Shing L. An improved tarpit for network deception[D]. Monterey: Naval Postgraduate School Monterey United States, 2016.

[31] Kim J, Shin S. Software-defined HoneyNet: Towards mitigating link flooding attacks[C]//47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops,2017:99 - 100.

[32] Achleitner S, Porta T F, Mcdaniel P, et al. Deceiving network reconnaissance using SDN-based virtual topologies[J]. IEEE Transactions on Network and Service Management, 2017,14(4):1098 - 1112.

[33] Wang H, Wu B. SDN-based hybrid honeypot for attack capture[C]//3rd Information Technology, Networking, Electronic and Automation Control Conference,2019:1602 - 1606.

[34] Zhang T, Lu B, Li D, et al. Anti-reconnaissance model of host fingerprint based on virtual node[C]//4th International Conference on Data Mining, Communications and Information Technology,2020:1584.

[35] Sun J H, Liu S, Sun K. A scalable high fidelity decoy framework against sophisticated cyber attacks[C]//6th ACM Workshop on Moving Target Defense,2019:37 - 46.

[36] Fan W J, Du Z H, Smith-Creasey M, et al. HoneyDOC: An efficient honeypot architecture enabling all-round design[J]. IEEE Journal on Selected Areas in Communications,2019,37(3):683 - 697.

[37] Lin Y D, Shih T B, Wu Y S, et al. Secure and transparent network traffic replay, redirect, and relay in a dynamic malware analysis environment[J]. Security and Communication Networks,2014,7(3):626 - 640.

[38] Hirata A, Miyamoto D, Nakayama M, et al. INTERCEPT +: SDN support for live migration-based honeypots[C]//4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security,2015:16 - 24.

[39] Araujo F, Hamlen K W, Biedermann S, et al. From patches to honey-patches: Lightweight attacker misdirection, deception, and disinformation[C]//ACM SIGSAC Conference on Computer and Communications Security,2014:942 - 953.

[40] Osman A, Bruckner P, Salah H, et al. Sandnet: Towards high quality of deception in container-based microservice architectures[C]//IEEE International Conference on Communications,2019:1 - 7.

[41] Wang Z Q, Li G F, Chi Y P, et al. Honeynet construction based on intrusion detection[C]//3rd International Conference on Computer Science and Application Engineering, 2019:1 - 5.

[42] Chuang P J, Hung T C. Enhanced attack blocking in IoT environments: Engaging honeypots and machine learning in SDN OpenFlow switches[J]. Journal of Applied Science and Engineering,2020,23(1):163 - 173.

[43] Urias V E, Stout W, Loverro C. Computer network deception as a moving target defense[C]//International Carnahan Conference on Security Technology,2015:1 - 6.

[44] Lin H, Zhuang J N, Hu Y C, et al. DefRec: Establishing physical function virtualization to disrupt reconnaissance of power grids' cyber-physical infrastructures[C]//Network and Distributed System Security Symposium,2020:23 - 26.

[45] McGrew R. Experiences with honeypot systems: Development, deployment, and analysis[C]//39th Annual Hawaii International Conference on System Sciences,2006:220.

[46] Al-Mohannadi H, Awan I, Hamar J. Analysis of adversary activities using cloud-based web services to enhance cyber threat intelligence[J]. Service Oriented Computing and Applications,2020,14(3):175 - 187.

[47] Kamel N, Eddabbah M, Lmoumen Y, et al. A smart agent design for cyber security based on honeypot and machine learning[J]. Security and Communication Networks,2020, 2020(1):8865474.

[48] Boukela L, Zhang G X, Bouzefrane S, et al. An outlier ensemble for unsupervised anomaly detection in honeypots data [J]. Intelligent Data Analysis,2020,24(4):743 - 758.

[49] Defibaugh-Chavez P, Veeraghattam R, Kannappa M, et al. Network based detection of virtual environments and low interaction honeypots[C]//IEEE Information Assurance Workshop,2006:283 - 289.

[50] Vetterl A, Clayton R. Bitter harvest: Systematically fingerprinting low-and medium-interaction honeypots at internet scale[C]//12th USENIX Workshop on Offensive Technologies,2018:1 - 9.

[51] Borders K, Falk L, Prakash A. OpenFire: Using deception to reduce network attacks[C]//3rd International Conference on Security and Privacy in Communications Networks and the Workshops,2007:224 - 233.

[52] Kyung S, Han W, Tiwari N, et al. HoneyProxy: Design and implementation of next-generation honeynet via SDN[C]// IEEE Conference on Communications and Network Security, 2017:1 - 9.

[53] Naik N, Jenkins P, Savage N. Threat-aware honeypot for discovering and predicting fingerprinting attacks using principal components analysis[C]//IEEE Symposium Series on Computational Intelligence,2018:623 - 630.

[54] Naik N, Shang C J, Shen Q, et al. Intelligent dynamic honeypot enabled by dynamic fuzzy rule interpolation[C]//20th International Conference on High Performance Computing and Communications,2018:1520 - 1527.

[55] Hayatle O, Otrok H, Youssef A. A Markov decision process model for high interaction honeypots[J]. Information Security Journal,2013,22(4):159 - 170.

[56] Dowling S, Schukat M, Barrett E. Using reinforcement learning to conceal honeypot functionality[C]//European Conference on Machine Learning and Knowledge Discovery in Databases,2019:341 - 355.