网络空间安全前沿技术丛书

野蛮人已经到来，烽火台烽燧已起，号角吹响战火已燃

顾众 沈卢斌 译

防范互联网上的“野蛮人”

网络钓鱼检测、DDoS防御和网络攻防实战

[美] 奥鲁瓦图比·艾约德吉·阿坎比（Oluwatobi Ayodeji Akanbi）
伊拉基·萨迪克·阿米里（Iraj Sadegh Amiri）
埃拉·费泽德科迪（Elahe Fazeldehkordi）
穆罕默德·雷扎·加利夫·索塔尼亚（Mohammad Reza Khalifeh Soltanian）
亨利·达尔齐尔（Henry Dalziel）
著

清華大學出版社

网络空间安全前沿技术丛书

防范互联网上的“野蛮人”

网络钓鱼检测、DDoS防御和网络攻防实战

奥鲁瓦图比·艾约德吉·阿坎比（Oluwatobi Ayodeji Akanbi）
伊拉基·萨迪克·阿米里（Iraj Sadegh Amiri）
[美] 埃拉·费泽德科迪（Elahe Fazeldehkordi） 著
穆罕默德·雷扎·加利夫·索塔尼亚（Mohammad Reza Khalifeh Soltanian）
亨利·达尔齐尔（Henry Dalziel）

顾众 沈卢斌 译

清华大学出版社
北京

图书在版编目(CIP)数据

防范互联网上的“野蛮人”：网络钓鱼检测、DDoS防御和网络攻防实战/(美)奥鲁瓦图比·艾约德吉·阿坎比等著；顾众，沈卢斌译. —北京：清华大学出版社，2019

(网络空间安全前沿技术丛书)

书名原文：A Machine Learning Approach to Phishing Detection and Defense；Theoretical and Experimental Methods for Defending Against DDOS Attacks；How to Attack and Defend Your Website

ISBN 978-7-302-51903-4

Ⅰ. ①防… Ⅱ. ①奥… ②顾… ③沈… Ⅲ. ①互联网络－网络安全－研究 Ⅳ. ①TP393.08

中国版本图书馆CIP数据核字(2018)第286720号

责任编辑：梁 颖 李 晔
封面设计：常雪影
责任校对：焦丽丽
责任印制：李红英

出版发行：清华大学出版社
网 址：http://www.tup.com.cn，http://www.wqbook.com
地 址：北京清华大学学研大厦A座 **邮 编**：100084
社 总 机：010-62770175 **邮 购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 装 者：三河市吉祥印务有限公司
经 销：全国新华书店
开 本：190mm×235mm **印 张**：11.75 **字 数**：277千字
版 次：2019年7月第1版 **印 次**：2019年7月第1次印刷
定 价：59.00元

产品编号：073557-01

顾众　电信系统从业者。十余年从事无线通信系统、电信平台产品、虚拟化及云计算的系统设计和研发，对嵌入式系统、网络技术、存储技术、虚拟化、信息安全等技术领域拥有丰富的经验和深刻的理解。目前为爱立信平台产品负责人，主要负责探索、规划产品未来的演进方向和组织产品新功能的开发及发布。

沈卢斌　毕业于瑞典皇家理工学院（KTH）片上系统设计（SoC）专业；长期从事电信云计算平台的研发和系统安全的测试工作；是最先同时通过CCIE和HCIE认证的网络专家之一；对SDN安全、FPGA在工业互联网交换机中的设计以及基于FPGA的入侵检测系统有浓厚的兴趣和独到的见解；目前主要从事SD-WAN产品架构设计和SD-WAN网络平台技术演进的研究工作。

译有《Google Hacking：渗透性测试者的利剑（原书第3版）》《信息安全分析学：大数据视角下安全的内核、模式和异常》《悄无声息的战场：无线网络威胁和移动安全隐私》和《防范互联网上的“野蛮人”：钓鱼检测、DDoS防御和网络攻防》等。

北京市版权局著作权合同登记号　图字：01-2017-5195
A Machine Learning Approach to Phishing Detection and Defense
Oluwatobi Ayodeji Akanbi, Iraj Sadegh Amiri, Elahe Fazeldehkordi
ISBN：978-0-12-802927-5

Theoretical and Experimental Methods for Defending Against DDOS Attacks
Mohammad Reza Khalifeh Soltanian, Iraj Sadegh Amiri
ISBN：978-0-12-805391-1

How to Attack and Defend Your Website
Henry Dalziel
ISBN：978-0-12-802732-5

Authorized Chinese translation published by Tsinghua University Press Ltd.

防范互联网上的“野蛮人”：网络钓鱼检测、DDoS 防御和网络攻防实战（顾众，沈卢斌　译）
ISBN：978-7-302-51903-4

注意

本书涉及领域的知识和实践标准在不断变化。新的研究和经验拓展我们的理解，因此须对研究方法、专业实践或医疗方法作出调整。从业者和研究人员必须始终依靠自身经验和知识来评估和使用本书中提到的所有信息、方法、化合物或本书中描述的实验。在使用这些信息或方法时，他们应注意自身和他人的安全，包括注意他们负有专业责任的当事人的安全。在法律允许的最大范围内，爱思唯尔、译文的原文作者、原文编辑及原文内容提供者均不对因产品责任、疏忽或其他人身或财产伤害及/或损失承担责任，亦不对由于使用或操作文中提到的方法、产品、说明或思想而导致的人身或财产伤害及/或损失承担责任。

本书献给

致力于，
商用SD-WAN智能广域网平台应用开发的
华斧网络科技（AXESDN）公司
所有网络专家

译者序

计算机网络最早出现于20世纪60年代。早期的计算机网络功能比较简单，主要供科研机构内部使用，不同机构之间的网络并没有大规模地互相连接。从20世纪80年代开始，计算机网络逐渐进入大量民用领域，政府机构、商业公司和其他实体之间的计算机通过跨越国家的网络连接起来，互联网作为一种新型的跨地域通信方式应运而生。随后而来的美国信息高速公路计划则引起了互联网在全球范围内的爆炸式发展，电子商务、网络支付和网络游戏等新型商业模式大量涌现，经过之后二十多年的成长和演化，互联网在今天已经渗透到我们生活中的方方面面。统计数据显示，2016年，中国的网络经济营业收入已经超过1万亿元，而且规模仍在高速增长中。

人们常说，有人的地方就有江湖，有江湖的地方就有战争。回顾人类社会的历史，文明的发展始终伴随着侵略与被侵略。从欧洲到中国，历史上的先进文明被野蛮文明毁灭的例子层出不穷。而每一次这类事件的后果都类似：大量的生命和财产被毁灭，社会的文明程度大幅倒退。所以一个文明如果要长期生存、发展和繁荣，必须拥有一套完善的防御体制来抵御野蛮人的进攻。

与人类文明的发展类似，在互联网诞生和发展的过程中，与之相伴而生的是形形色色的互联网攻击。早期的攻击者大多是为了炫耀自己在计算机和网络技术方面的超人造诣，并没有涉及太多经济利益。随着互联网逐渐进入到社会经济活动中，网络攻击逐渐转向了以谋取经济利益为目的。近些年来，电子商务和互联网金融的普及使得网络攻击的规模变得更为庞大，造成的经济和其他方面的损失也更加严重。根据美国一些机构的统计，网络攻击每年对美国造成多达数百亿甚至上千亿美元的经济损失。例如，在2016年发生的针对国际支付系统(Society for Worldwide Interbank Financial Telecommunications，SWIFT)的攻击中，孟加拉国银行被窃取了高达8100万美元的资产。网络攻击的受害者不仅仅是大公司，针对个人身份及其他私密信息的犯罪则威胁着我们每个人的安全。统计数据显示，仅在2016年美国就有1500万名公民成为身份盗窃的受害者。

因而，如何防范互联网上的“野蛮人”，也就是那些使用形形色色手段发起攻击的网络犯

罪者，是今天这个网络时代每一个公司，甚至每个人都需要认真关注的问题。然而，就像武林中的种种功夫流派一样，网络攻击的手段种类繁多。可惜的是，网络世界至今尚未出现“九阳神功”这样的无上秘籍，能够一劳永逸地防御所有类型的网络攻击。因此，针对每类特定的攻击方式，需要研究其工作原理及弱点来构建特定的防御手段进行反制。在研究和学习网络攻防的过程中，我们阅读了相关的大量文献，在此过程中，有幸研读了关于拒绝服务攻击和网络钓鱼防御的两本专著。考虑到这两类攻击发生的频率非常之高，而且造成的经济损失和其他后果相当严重，我们深感如果能够把这些专著介绍给国内的读者，会有益于网络用户和网络安全专业人员了解这两类攻击方式的原理、特点和常用防御方式。此外，我们简单介绍了使用一种开源软件破解网站的基本步骤以及破解过程中利用到的系统漏洞。所谓知己知彼、百战不殆，网络安全专业人员可以通过这个破解软件来检测网站的漏洞以及潜在的风险，来改进并提升网站的安全性。

通过机器学习来防御钓鱼攻击的专著格外引起了我们的兴趣。其实作为人工智能的核心技术，机器学习算法并非新生事物，但是直到近十年来，随着计算机硬件性能的提升、算法的改进以及海量数据的有效利用，才使得这些算法能够大量被应用在各种各样的现实场景中。尤其是近两年自动驾驶和其他各种机器人的大量涌现，更是让人惊呼人工智能主导的下一次工业革命即将到来。我们相信这本专著不但可以帮助读者了解如何利用机器学习来防御网络钓鱼，而且这些知识也有助于读者适应并投身于这场由人工智能引起的巨大变革中。

本书的内容组织为三部分：第一部分专注于网络钓鱼的工作原理及如何应用机器学习算法来防御网络钓鱼；第二部分则讨论了分布式拒绝服务攻击和防御的方法；第三部分介绍了攻击网站服务器的基本原理，以及使用开源软件 Burp Suite 发动攻击的具体步骤。

我们很荣幸有机会把这些研究介绍给国内的专业人士，在翻译过程中，我们秉持尊重原著的理念，尽量在中文译本中保持原意。但是在保持原意的基础上，做了一定的修改以更加符合中文的阅读习惯。由于译者水平所限，书中难免有不准确或不精确之处，敬请读者不吝指正。

在本书的翻译过程中，我们得到了清华大学出版社相关人员的大力帮助，在此表示诚挚的感谢。

译　者

2018 年 7 月于上海

目 录

第一篇 机器学习方法检测钓鱼网站

第二篇 分布式拒绝服务攻击防御实践

第一篇
机器学习方法检测钓鱼网站

网络钓鱼是一种利用欺骗性电子邮件和假冒网站窃取用户个人信息的网络攻击。本篇首先回顾钓鱼检测领域的相关研究工作，然后描述一种检测钓鱼网站的新型组合算法及相关实验工作。实验工作主要由三个阶段组成：第一个阶段关注数据采集、预处理、特征提取和数据分割；第二阶段对四种分类算法(C4.5、SVM、*K*-NN 和 LR)在精确率、召回率、准确率和 *f* 值等方面做性能评估，并找出性能最佳的单个分类算法；最后阶段评估多种组合算法的性能并找出最佳组合分类算法，与最佳单个分类算法进行性能对比。结果表明，*K*-NN 分类算法达到了 99.37%的准确率，而组合分类的准确率为 99.31%。其原因是实验中使用了较小型的数据集，而 *K*-NN 算法本身恰恰更适合处理小型数据集。实验中用到的另两个分类算法(SVM 和 C4.5)则更适合处理大型数据集。

缩写表

ANN Artificial Neural Network 人工神经网络

APWG Anti-Phishing Work Group 反网络钓鱼工作组

BART Bayesian Additive Regression Trees 贝叶斯累加回归树

C4.5 Decision Tree 决策树

CA Certificate Authority 认证授权

DNS Domain Name System 域名系统

DR Detection Rate 检测率

ENS Ensemble 组合

HTML Hyper Text Markup Language 超文本标记语言

HTTP Hyper Text Transfer Protocol 超文本传输协议

HTTPS Hyper Text Transfer Protocol Secure 超文本传输安全协议

IP Internet Protocol 因特网协议

***K*-NN** *K*-Nearest Neighbor *K*-最近邻(算法)

LR Linear Regression 线性回归(算法)

MLP Multi-Layer Perceptron 多层感知器

NB Naäve Bayesian 朴素贝叶斯

Pred. Prediction 预测

ROC Receiver Operating Characteristic

受试者操作特征(曲线)(ROC 曲线)

SQL Structured Query Language 结构化查询语言

SSL Secure Socket Layer 安全套接层

SVM Support Vector Machine 支持向量机(算法)

TTL Time to Live 生存时间

URL Uniform Resource Locator 统一资源定位符

URI Uniform Resource Identifier 统一资源标识符

FP False Positive 虚警(预测为正,实际为负)

FN False Negative 漏警(预测为负,实际为正)

TP True Positive 真阳性(预测为正,实际为正)

TN True Negative 真阴性(预测为负,实际为负)

FPR False Positive Rate 虚警率,同 FAR(False Alarm Rate)

FNR False Negative Rate 漏警率

TPR True Positive Rate 召回率,查全率,同 Recall

第1章
背景介绍

摘要

本章内容大体组织如下：首先，简单介绍网络钓鱼的基本概念和目前常见的钓鱼攻击防御技术；其次，简单回顾钓鱼攻击的历史并且解释为何它会成为网络安全领域的一个重点研究课题，在这部分也会讨论钓鱼攻击对电子商务的影响；最后简单介绍研究方法、期望的研究结果和本研究的重要性，以及对未来研究工作的展望。

关键词

网络钓鱼
网络安全
网站
信息
威胁
分类算法
漏警

1.1 绪论

网络犯罪是指针对计算机或网络的犯罪[Martin et al., 2011]，包括多种潜在的犯罪行为。根据网络犯罪针对的目标，它们可以分为两个主要类别：

(1) 第一类是针对计算机、网络或其他计算设备的犯罪；

(2) 第二类则是以计算机、网络或其他电子设备为工具，而非以此为目标的犯罪。

网络钓鱼是一种相对较新的网络犯罪活动，它的目标并非针对计算机本身，而是通过网络窃取他人的身份信息或者其他个人隐私。钓鱼攻击的主要危害是钓鱼者在窃取了受害者的私密信息之后，通过滥用这些信息窃取受害者的财产或者其他贵重物品。相较于黑客和病毒等其他形式的网络威胁，钓鱼攻击是一种快速增长的网络犯罪活动。

由于互联网已经成为现代社会的主要通信方式，钓鱼攻击的发起方式通常有以下几种[Alnajim & Munro, 2009]。

- 电子邮件到电子邮件：攻击者发送欺诈电子邮件要求收信者发送敏感信息给攻击者。
- 电子邮件到网站：攻击者发送给受害者包含钓鱼网站地址的电子邮件。
- 网站到网站：受害者单击搜索引擎或在线广告上的钓鱼网站链接。
- 浏览器到网站：受害者在浏览器上输错网址打开与合法网址非常相似的钓鱼网站。

网络钓鱼的攻击者通常使用诈骗邮件和虚假网站来引诱客户泄露银行账户信息，网站登录信息等个人机密[Topkara et al., 2005]。常见的一种做法是，网络钓鱼者向用户发送带有网站重定向链接的电子邮件要求用户更新机密信息，如合法的信用卡信息、网站登录信息和银行账户信息等。当用户单击其中的链接时，打开的其实是钓鱼者设立的山寨网站，用户在山寨网站输入的机密信息将会被钓鱼者截获。由于克隆一个银行或者其他涉及用户隐私的网站极其容易，所以封堵网络钓鱼非常困难。正如文献[Aburrous et al., 2008]解释的那样，理解和分析钓鱼攻击的难处在于这种欺诈方式涉及的技术和人性的复杂度。

当前有多种类型的反网络钓鱼措施可用于防范钓鱼攻击。例如，反网络钓鱼工作组是一个行业组织，它从不同来源收集关于网络钓鱼的信息，编辑成相关报告提供给其付费成员[RSA, 2006]。现今的浏览器也通过扩展程序或者工具栏把反钓鱼应对措施嵌入到网站登录操作中。许多浏览器的工具栏提供了可用于检测网络钓鱼行为的功能。文献[Garera et al., 2007]提出的 SpoofGuard 程序，给我们带来了利用网址、图像、域名和链接来评估欺诈的可能性，因为检测到钓鱼网站时它会向用户发出警告[Chou et al., 2004]。

朗讯个性化网络助手(Lucent Personalized Web Assistant, LPWA)是一种防止身份被盗窃的个人信息保护工具[Gabber et al., 1999; Kristol et al., 1998]。它使用一个函数来定义用户的相关变量，如用户访问每个服务器的电子邮件地址、用户名和密码等。文献[Ross et al., 2005]提出的 PwdHash 软件使用了类似的方法。

在一个称为人类交互验证的试验中，文献[Dhamija & Tygar, 2005a]提出了通过人力来区分合法网站和欺诈网站之间的特征。在这个工作的基础上，文献[Dhamija & Tygar,