网络钓鱼欺诈检测技术研究

张茜[1,2,3]，延志伟[3]，李洪涛[3]，耿光刚[3]
（1. 中国科学院计算机网络信息中心，北京 100190；
2. 中国科学院大学，北京 100049；
3. 中国互联网络信息中心互联网域名管理技术国家工程实验室，北京 100190）

摘　要：分析了网络钓鱼欺诈的现状，并对钓鱼检测常用的数据集和评估指标进行了总结。在此基础上，综述了网络钓鱼检测方法，包括黑名单策略、启发式方法、视觉匹配方法、基于机器学习的方法和基于自然语言理解的方法等，对比分析了各类方法的优缺点，进一步指出了钓鱼检测面临的挑战，并展望了钓鱼检测未来的研究趋势。

关键词：网络钓鱼欺诈；钓鱼检测；机器学习；视觉匹配
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.2096-109x.2017.00180

Research of phishing detection technology

ZHANG Xi[1,2,3], YAN Zhi-wei[3], LI Hong-tao[3], GENG Guang-gang[3]
(1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. National Engineering Laboratory for Internet Domain Name Management, China Internet Network Information Center, Beijing 100190, China)

Abstract: The current status of phishing scams were analyzed and the data sets and evaluation indicators commonly used in phishing detection were summaried. On this basis, a detailed overview of the typical methods of phishing detection was given, which included blacklist strategies, heuristic methods, visual matching methods, and methods based on machine learning and natural language processing. The comparison and analysis of those methods were given, and furtherly, the challenges and future trends of phishing detection were discussed.

Key words: phishing fraud, phishing detection, machine learning,visual matching

1　引言

国家互联网信息办公室于 2016 年 12 月 27 日发布的《国家网络空间安全战略》指出，要严厉打击网络诈骗、网络盗窃等违法犯罪行为[1]。随着互联网的发展，互联网犯罪事件频有发生，严重损害了国家、企业和个人利益。网络钓鱼是实施网络诈骗、网络盗窃的主要手段，对网络钓鱼的检测已成为网络空间安全研究中的一个重要领域。

网络钓鱼（phishing）这一术语产生于 1996 年，它是由钓鱼（fishing）一词演变而来。在网络钓鱼的过程中，攻击者使用诱饵（如电子邮件、手机短信）发送给大量用户，期待少数用户“上钩”，进而达到“钓鱼”（如窃取用户的隐私信息）的目的。国际反网络钓鱼工作组（APWG, Anti-

收稿日期：2017-06-13；**修回日期**：2017-07-05。**通信作者**：耿光刚，gengguanggang@cnnic.cn
基金项目：国家自然科学基金资助项目（No.61375039）
Foundation Item: The National Natural Science Foundation of China (No.61375039)

Phishing Working Group）给网络钓鱼的定义是：网络钓鱼是一种利用社会工程学和技术手段窃取消费者的个人身份数据和财务账户凭证的网络攻击方式[2]。采用社会工程手段的网络钓鱼攻击往往是向用户发送貌似来自合法企业或机构的欺骗性电子邮件、手机短信等，引诱用户回复个人敏感信息或单击里面的链接访问伪造的网站，进而泄露凭证信息（如用户名、密码）或下载恶意软件。而技术手段的攻击则是直接在 PC 上移植恶意软件（如浏览器中间者（MitB, man-in-the-browser）攻击），采用某些技术手段直接窃取凭证信息，如使用系统拦截用户的用户名和密码、误导用户访问伪造的网站等。

攻击者实施网络钓鱼攻击的重要目的有以下两点[3]。

1) 获取经济利益：攻击者通过将窃取到的身份信息卖出或者直接使用窃取到的银行账户信息获得经济利益。

2) 展示个人能力：网络钓鱼攻击者为了获得同行的认同而实施网络钓鱼活动。

近年来，网络钓鱼攻击已经成为互联网用户、组织机构、服务提供商所面临的最严重的威胁之一。据易安信公司信息安全事业部（RSA）估计，2014 年 12 月，全球的组织机构由于网络钓鱼所遭受的经济损失约 4.53 亿美元[4]。中国反钓鱼联盟（anti-phishing alliance of China）的报告也指出，网民一年之内因网络欺诈的损失高达 300 多亿元，30%的网购者曾遭遇钓鱼网站的攻击[5]。尽管目前已经有多种反钓鱼工具和技术用来遏制钓鱼攻击，网络钓鱼的数量依然增长迅速。国际反网络钓鱼工作组 2016 年的统计报告显示，2016 年第二季度共检测到钓鱼网站 466 065 个，与 2015 年第四季度相比，增加了 61%[2]。图 1 显示的是 2014~2016 年各季度 APWG 所检测到的钓鱼网站的数目[注1]，从图 1 中可以看出，2014 年以来，虽然有所波动，但钓鱼网站的数量整体呈持续增长的趋势。国内方面，截至 2016 年 9 月，APAC 累计认定并处理钓鱼网站 382 092 个，其中仅 2016 年上半年就处理了 79 719 个钓鱼网站，远超 2015 年全年的数量（58 660 个）[6,7]。网络钓鱼的日益猖獗使互联网用户面临身份欺诈、个人隐私信息泄露以及经济损失等各方面的威胁。因此，如何有效地检测并处理网络钓鱼已成为亟待解决的网络安全问题。

网络钓鱼发展至今，其针对的目标已经从互联网终端用户扩展到了组织机构、网络提供商，也有了更为复杂的网络钓鱼形式，如近年来愈加严峻的鱼叉式网络钓鱼攻击（spear phishing）。在鱼叉式网络钓鱼中，攻击者通常会锁定特定个人或某机构的特定员工及其社交账号，向其发送个性化的电子邮件，诱使他们泄露敏感信息或在电脑上安装恶意软件。尽

注1　数据来自 APWG 发布的报告。

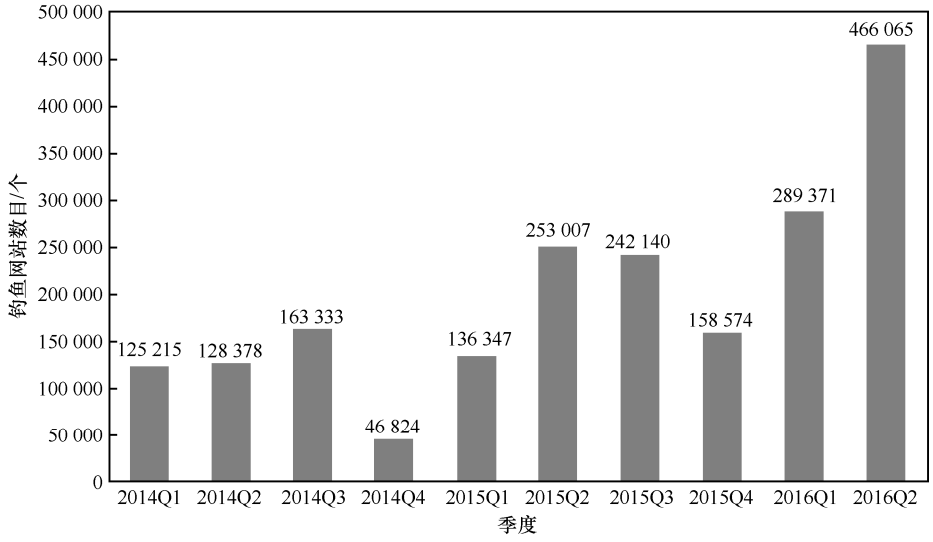


图 1　2014~2016 年各季度 APWG 检测到的钓鱼网站数目

管鱼叉式网络钓鱼只是发送少量的邮件给少量的目标，但个性化的特点使其与一般的网络钓鱼相比，更难以检测且具有更高的成功率[8,9]。FBI 指出，一种名为“执行长欺诈”（CEO fraud）的钓鱼在 2013 年 10 月到 2016 年 2 月期间造成的损失高达 23 亿美元[10]。

钓鱼检测技术通过利用钓鱼攻击所具有的某些特征对其进行识别，从而实现对网络钓鱼攻击的打击和防范。本文统计了 2006~2015 年网络钓鱼检测相关专利、文献的发表数目[注2]，如图 2 所示，钓鱼检测相关研究成果的数目整体呈上升趋势。

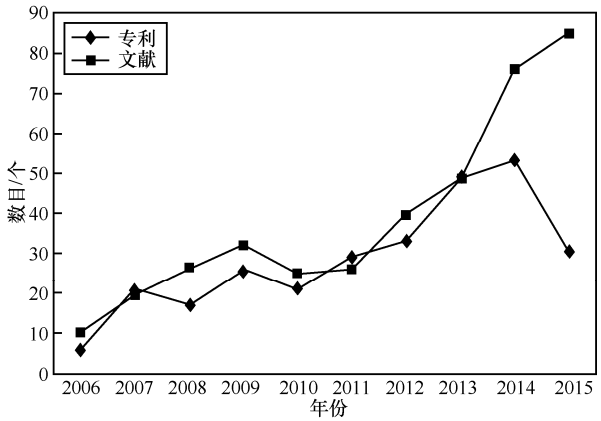


图 2　2006~2015 年钓鱼检测相关专利、文献发表数目

国内目前钓鱼检测的相关研究很多，但缺乏论述全面、条理清晰的综述性文献。因此，本文尝试对网络钓鱼检测的思路、方法、技术进行全面的归纳和总结。

2　网络钓鱼检测视角分析、语料库及评价指标

2.1　钓鱼检测视角分析

网络钓鱼的攻击和防御就像一场持续的“军备竞赛”，尽管目前已有许多关于钓鱼检测的技术研究和实现，但它们无法有效防御所有的网络钓鱼攻击。一方面，网络钓鱼攻击者常会根据已有的钓鱼检测方案改进钓鱼策略，达到规避检测的目的；另一方面，网络钓鱼活动具有伪装性高、时效性强、存活时间短及钓鱼目标广泛等特点[11]，往往很难有效地识别。

虽然网络钓鱼的模式在不断地演化，但其本质并未发生变化。网络钓鱼总是与其仿冒的目标有很强的关系，并存在一定的迷惑性信息。例如与合法链接相似的域名、使用指向合法页面的链接以及视觉上相似的内容等，才能诱导用户输入自己的敏感信息。网络钓鱼检测就是发现并利用这些与合法内容（URL、邮件、网页等）有关的迷惑性信息进行网络钓鱼的检测和识别的。

网络钓鱼攻击者进行网络钓鱼的流程如图 3 所示。首先，攻击者假设一个钓鱼网站或使合法网站携带恶意代码，并部署一些必需的后台脚本用于处理并获取用户的输入数据。然后，攻击者利用社会工程学[注3]制作诱饵，并通过邮件、电话、短信等途径发放诱饵。在用户被引诱访问钓鱼页面并上传隐私信息后，攻击者即可利用事先实现的后台程序得到这些信息，并利用用户隐私信息牟取利益。

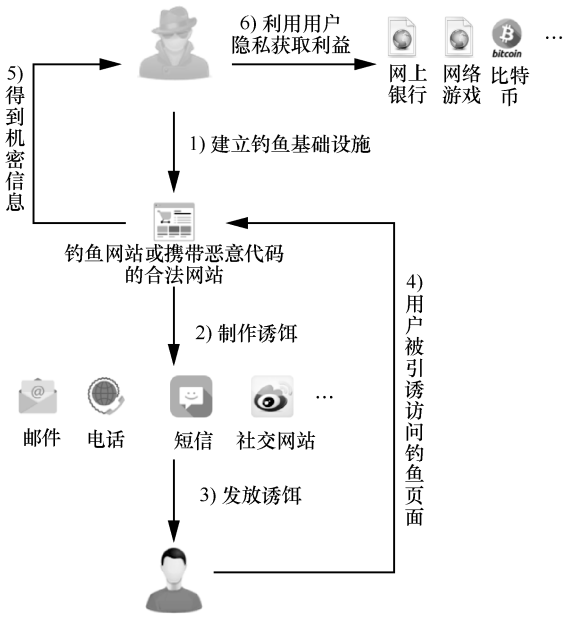


图 3　网络钓鱼攻击流程

目前常用的网络钓鱼检测方法的分类方式有很多，从检测的视角来看，根据所关注的钓鱼攻

注 2　数据来自 Web of science 检索结果。

注3 攻击者利用“人”自身的弱点（往往是心理学层面）来获取信息、影响他人，从而达到不可告人的目的。

击的不同实施阶段——钓鱼攻击的发起从图 3 中的阶段 3 发放诱饵开始，钓鱼检测的方法可以分为：基于传播途径分析的方法、基于网站入口分析的方法和基于网站内容分析的方法。根据检测手段又可以分为基于黑名单的钓鱼检测、启发式钓鱼检测、基于视觉相似性的钓鱼检测、基于机器学习的钓鱼检测以及基于自然语言处理技术的钓鱼检测（将在第 3 部分详细介绍）。这 2 种分类方式之间相互交叉，图 4 简明地描述了两者之间的关系，其中方块颜色的深浅表示使用频率的高低。

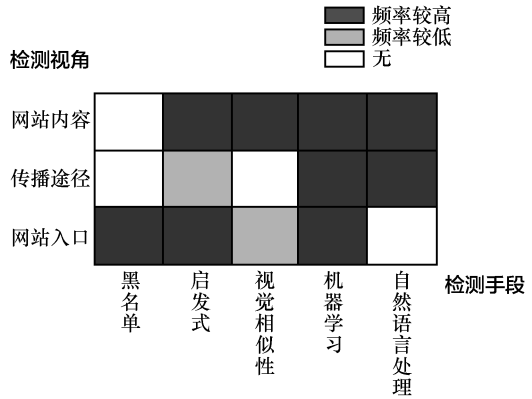


图 4 检测手段与检测视角之间的关系

2.2 基于传播途径分析的方法

网络钓鱼的传播途径包括电子邮件、短信、电话、即时信息、各种社交平台（微博、Twitter 等）及其他新的通信方式。网络钓鱼信息的传播和扩散是攻击者发动钓鱼攻击的第一个阶段，在这一阶段进行网络钓鱼的检测可以将钓鱼信息直接过滤，使其无法到达终端用户，从而构成钓鱼攻击的第一道防线。目前有关传播途径的钓鱼检测研究中对短信钓鱼（Smishing，SMS phishing）检测[12]、电话钓鱼（Vishing，voice phishing）[13,14]检测等的研究并不多，主要关注的是电子邮件钓鱼检测[15~19]。

电子邮件钓鱼检测通过对用户收到的电子邮件进行分析，对邮件中是否包含钓鱼信息进行判断、过滤。钓鱼邮件一般有 2 种情况：一是包含钓鱼网站链接，引诱用户去访问；二是不包含任何链接，而是利用用户的好奇心，诱导他们回复敏感信息[17]。图 5 概括了基于电子邮件分析的方法中常用的特征。

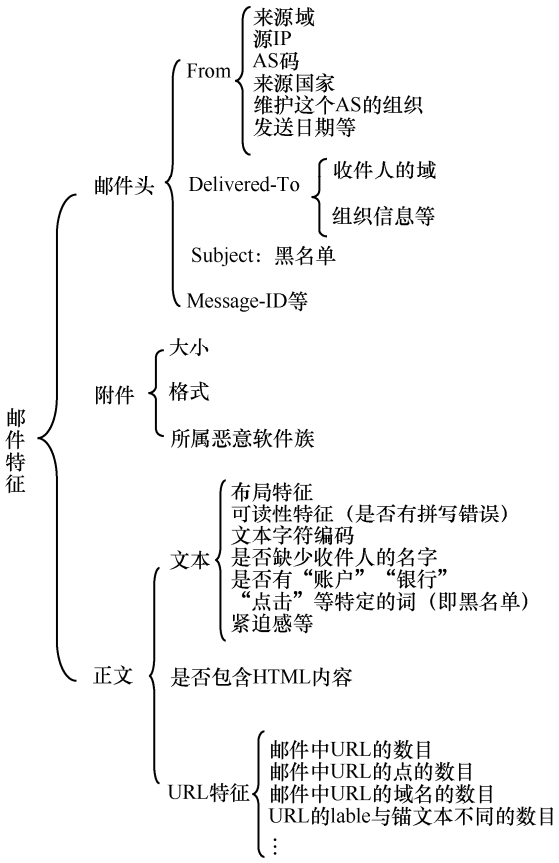


图 5 电子邮件常用特征

一封电子邮件主要包含 3 部分：邮件头、正文、附件。邮件头由多个预先定义的格式化字段组成，如 From、Delivered-To、Subject、Message-ID[20]等。网络钓鱼攻击者虽然可以将邮件伪装成来自合法的组织或机构，却无法隐藏电子邮件的真实来源、Message-ID 等信息。电子邮件的正文部分是邮件的主要内容，通常是 Text 或 HTML 格式的。钓鱼邮件的正文有很多特征。例如，称呼只使用统称而非收件人的名字、刻意营造紧迫感（如要求用户立即更新账户信息，否则会有账户被盗的风险）及可疑的统一资源定位符（URL，uniform/universal resource locator）等，是钓鱼邮件检测的主要特征来源。此外，钓鱼邮件的附件中往往包含侦察软件或木马病毒，因此确认邮件附件的合法性是钓鱼邮件检测中必不可少的一环。

2.3 基于网站入口分析的方法

URL 是因特网上标准的资源地址，即网站的入口。URL 仿冒在网络钓鱼中很常见，引诱用户单击 URL 访问其搭建的钓鱼网站是网络钓鱼的

重要环节之一。为了提高用户访问钓鱼网站的可能性，钓鱼攻击者往往使用与所仿冒的目标视觉上相似的、具有迷惑性的 URL。一个标准 URL 的格式如下。

protocol://hostname[:port]/path/[;parameters][?query] #fragment

常见的 URL 仿冒的方法是在目标 URL 的基础上对主机名[注4]（host name）部分和路径[注5]（path）部分进行部分修改替换来构造钓鱼 URL，以达到混淆视听的目的。例如，攻击者使用“www.lcbc.com.cn”仿冒工商银行（真实 URL 为“www.icbc.com.cn”），使用“www.cmb955555.com”仿冒招行网站（真实 URL“www.cmbchina.com”）等。

除了视觉上的相似性之外，钓鱼 URL 还具有许多其他特征。在网络钓鱼检测中常用的 URL 特征主要是词汇特征[21~26]和基于主机的特征[23~26]，如图 6 所示。

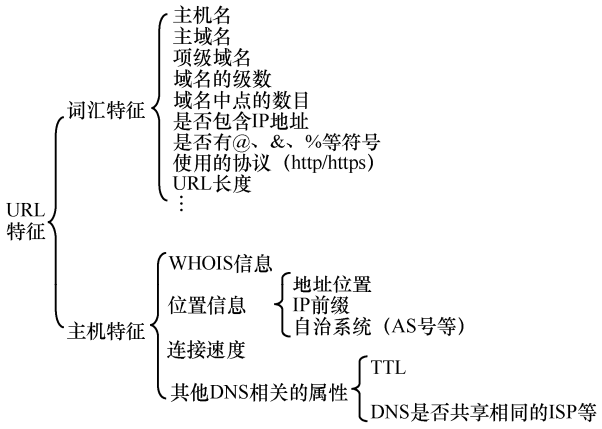


图 6　URL 常用特征

URL 的词汇特征是直接从 URL 中提取的特征，常使用“/”“?”“.”“=”“_”“&”和“-”作为分隔符，然后使用词袋模型对各词块进行表示。词汇特征能很好地捕捉钓鱼 URL 所具有的特点，如与合法域名相似，常包含@ 、&、%等特殊符号。

主机特征描述了 URL 主机名部分所标识的网站主机的属性，通过这些属性可以估计该钓鱼 URL 的位置、拥有者等信息。常用的主机特征一般有 WHOIS [注6]信息、位置信息、连接速度及其他 DNS 相关的属性等。

对 URL 进行分析在网络钓鱼检测的相关研究工作中使用率相当高，在基于传播途径分析的方法[16,18]和基于网站内容分析的方法[27,28]中都会用到。另外，URL 还是黑名单技术的主要对象[29]。但由于 URL 中并不具有钓鱼网站的决定性特征，即窃取用户信息的手段，具有局限性[30]，现在已很少有人进行单纯分析 URL 的研究。

2.4　基于网站内容分析的方法

钓鱼网页往往采用社会工程学手段的网络钓鱼攻击的最后一步，绝大多数的网络钓鱼最终都引诱用户访问其事先搭建好的仿冒网站。在这种情况下，基于网站内容分析的网络钓鱼检测实际上是反钓鱼的最后一道防线。

为了更好地取得用户的信任，钓鱼攻击者构建的钓鱼网页往往与真实网页十分相似，这种相似性包括 Logo 的相似性[31~33]、Favicon 的相似性[32,34]、CSS 架构的相似性[35,36]、布局的相似性[37~40]及网页整体视觉的相似性[37,41,42]，利用这种相似性及钓鱼网页与真实网页的不同之处进行目标品牌的识别和网络钓鱼的检测十分有效。

此外，对网站内容的分析还包括对网页底层 HTML [注7]的分析[27,43~45]。在网页的 HTML 中存在着许多有辨识性的特征，如标题、链出的 URL 与本网页 URL 的域名是否一致、URL 与其标签是否一致，是否有隐藏字段，是否有 Form 表单等。图 7 总结了基于网页内容分析方法中常用的特征。在有些研究中只使用了 HTML 的文本内容，通过 TF-IDF 算法得到整个页面的关键词[43,44,46]。但多数研究在对网站内容进行分析的时候会同时使用多种 HTML 特征，例如，文献[45]使用的 HTML 特征为是否包含有效的网络内容服务商

注4 存放资源的服务器的域名系统（DNS）主机名或 IP 地址。

注5 由零或多个“/”隔开的字符串，一般用来表示主机上的一个目录或文件地址。

注6 WHOIS 是用来查询域名的 IP 以及所有者等信息的传输协议。

注7 超文本标记语言，是标准通用标记语言下的一个应用，它通过标记符号来标记要显示的网页中的各个部分。

（ICP, internet content provider）、空链的数目、出链的数目及是否包含有效的电子商务证书信息；文献[27]中则使用了标题、文本、出链和版权声明这 4 个特征。

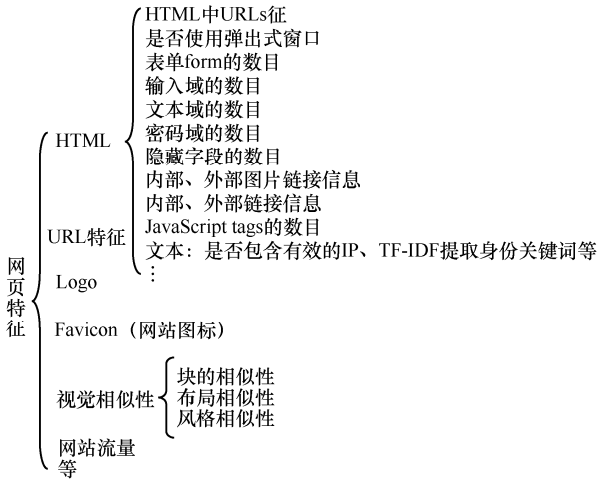


图 7 网页常用特征

每类特征都具有一定的针对性，在实际应用中，往往会将多类特征融合，从而尽可能地提高钓鱼检测的效果。例如，Zhang 等[47]融合使用了 URL 特征、文本特征及基于规则的特征；胡向东等[33]则使用了敏感文本特征和 Logo 图像特征进行金融类钓鱼网页的检测，具有很强的针对性和时效性；徐欢潇等[48]针对钓鱼网站有的以文字为主、有的以图片为主的现象，融合使用了文本特征、页面布局特征及 URL 链接特征。

2.5 常用语料库

在进行钓鱼检测的研究时，往往需要大量的网络钓鱼数据和合法数据（邮件、URL、网页 HTML、网页截图等），本文总结了一些常用的语料库。

PhishTank：PhishTank[49]是一个可以让用户提交、验证和共享网络钓鱼链接的社区网站。用户提交可疑的钓鱼 URL 后，会有至少 2 名网站成员进行人工检查。一旦确认为网络钓鱼，就会将该 URL 加到一个可供他人下载的数据库中。

Millersmiles：Millersmiles[50]是关于欺诈类电子邮件和网络钓鱼行为信息的重要信息来源，它包含了大量来自实际事例中与电子邮件、伪造的网页内容相关的文字类和图片类资料。

SpamAssassin public corpus：SpamAssassin[51]是一个旨在检测垃圾邮件和钓鱼邮件的免费开源软件项目，它的公共语料库中包含大量垃圾邮件和非垃圾邮件语料信息，可为网络钓鱼邮件的检测提供数据集。

MalwarePatrol：MalwarePatrol[52]是一个由用户贡献的免费系统。与 PhishTank 类似，任何人都可以提交可能携带恶意软件、病毒或木马的可疑网址。提交的 URL 被 MalwarePatrol 确认为恶意的之后，该 URL 就会被放入一个黑名单中，供用户下载。

Open Directory：开放目录专案[53]（即 DMOZ）是一个大型公共网页目录，它是由来自世界各地的志愿者共同维护和建设的全球最大目录社区[54]。这个目录下的网页依照其性质和内容分门别类，在进行钓鱼检测的研究时可以从中获取合法 URL 的数据集。

2.6 评价指标

网络钓鱼检测的目标是从包含了网络钓鱼实例和合法实例的数据集中检测出钓鱼实例，本质上是一个二分类问题。在二分类问题中，共有 4 种分类情况，常用混淆矩阵衡量分类的准确性（如表 1 所示）。其中，$N_{p\to p}$ 表示将钓鱼实例正确预测为钓鱼的数目，$N_{p\to l}$ 表示将钓鱼实例错误地预测为合法实例的数目，$N_{l\to p}$ 表示将合法实例错误地预测为钓鱼实例的数目，$N_{l\to l}$ 表示将合法实例正确预测为合法实例的数目。

表 1 混淆矩阵

实际实例类型	预测为钓鱼实例的数目	预测为合法实例的数目
实际为钓鱼实例	$N_{p\to p}$	$N_{p\to l}$
实际为合法实例	$N_{l\to p}$	$N_{l\to l}$

在网络钓鱼检测技术中，常用的性能评估指标如下。

1) 灵敏度（sensitivity）：将钓鱼实例预测为钓鱼实例的能力，见式(1)。

2) 特异度（specificity）：将合法实例预测为合法实例的能力，见式(2)。

3) 误检率（FPR, false positive rate）：将合法实例错误地预测为钓鱼实例的比例，见式(3)。

4) 漏检率（FNR, false negative rate）：将钓鱼实例错误地预测为合法实例的比例，见式(4)。

5) 准确率（P, prediction）：在所有预测为钓鱼的实例中，确实是钓鱼的实例所占的比例，见式(5)。

6) 召回率（R, recall）：等价于 sensitivity，见式(6)。

7) F-measure：准确率 P 和召回率 R 的加权调和平均数，计算如式(7)。其中 β 是参数，当 β=1 时，就是常见的 F_1 值，见式(8)。

9) 精确度（ACC, accuracy）：钓鱼实例和合法实例正确预测的比例，见式(9)。

9) 加权错误率（W_{Err}）：钓鱼实例和合法实例预测错误的加权错误率[55]，见式(10)。其中，λ 是权重系数，表示合法实例的重要程度。例如，若 λ=1，则钓鱼实例和合法实例的重要程度相同；若 λ=5，则对于将合法实例误检为钓鱼实例的惩罚是钓鱼实例漏检测惩罚的 5 倍。

$$sensitivity = \frac{N_{p\to p}}{N_{p\to p} + N_{p\to l}} \tag{1}$$

$$specificity = \frac{N_{l\to l}}{N_{l\to p} + N_{l\to l}} \tag{2}$$

$$FPR = \frac{N_{l\to p}}{N_{l\to p} + N_{l\to l}} \tag{3}$$

$$FNR = \frac{N_{p\to l}}{N_{p\to p} + N_{p\to l}} \tag{4}$$

$$P = \frac{N_{p\to p}}{N_{p\to p} + N_{l\to p}} \tag{5}$$

$$R = \frac{N_{p\to p}}{N_{p\to p} + N_{p\to l}} \tag{6}$$

$$F_\beta = \frac{(1+\beta^2)PR}{\beta^2 P + R} \tag{7}$$

$$F_1 = \frac{2PR}{P+R} \tag{8}$$

$$ACC = \frac{N_{p\to p} + N_{l\to l}}{N_{p\to p} + N_{p\to l} + N_{l\to p} + N_{l\to l}} \tag{9}$$

$$W_{Err} = 1 - \frac{N_{p\to p} + \lambda N_{l\to l}}{N_{p\to p} + N_{p\to l} + \lambda N_{l\to p} + \lambda N_{l\to l}} \tag{10}$$

3　网络钓鱼检测技术

3.1　基于黑名单的钓鱼检测

基于黑名单的检测方法维护一个已知的钓鱼网站的信息列表，以便根据列表检查当前访问的网站。这份需要不断更新的黑名单中包含已知网络钓鱼的 URL （如 PhishTank[49]）、IP 地址（如 spamhaus[56]）、域名（如 SURBL[57]）、证书（如证书撤销列表 CRLs [注8]）或者关键词等信息。

黑名单的方法应用广泛，是主要的网络钓鱼过滤技术之一，如 Google Chrome、Mozilla Firefox 和 Apple Safari 中使用的 Google Safe API[58]，就是根据 Google 提供的不断更新的黑名单，通过验证某一 URL 是否在黑名单中，来判断该 URL 是否是钓鱼网页或者恶意网页。

如何将可疑 URL 与黑名单中的网络钓鱼 URL 进行匹配是基于黑名单的方法中一个关键问题。为了规避黑名单的检测，网络钓鱼攻击者往往会不断改变钓鱼页面的 URL，而 URL 的任何一点变化都会导致与黑名单中的 URL 匹配失败，从而导致漏检情况的发生。针对精确匹配的局限性，Prakash 等[59]提出了一种改进方法 PhishNet，基于 5 种启发式的规则（如通用顶级域名的可替换性、目录结构相似性等）枚举已知网络钓鱼的简单组合，在经过 DNS 查询和页面内容匹配验证之后得到新的钓鱼 URL，然后将 URL 分解为 4 个部分——IP 地址、主机名称、目录结构和品牌名字，与黑名单中的相应部分进行近似匹配以判断 URL 是否是网络钓鱼。PhishNet 可以对黑名单列表进行扩充，并能检测出一部分未在黑名单中出现的网络钓鱼。

Felegyhazi 等[60]探讨了基于域名黑名单的主动型方法。该方法基于网络犯罪分子需要注册大量的域名以维持其活动这一发现，将一个域名黑名单作为种子列表，利用 DNS 区域文件（zone file）的 NS 信息和 WHOIS 域名注册信息对列表进行扩充。同时，该方法还利用名称服务器注册的新鲜度和自我解析等特征。结果表明，与以往被动的黑名单加入方式相比，这种主动将域名列

注8 证书撤销列表是在其计划的到期日期前被证书颁发机构（CA）撤销并且不再受到信任的数字证书的列表。

入黑名单的方法可以减少 60%～75%域名加入黑名单的时间间隔。但该方法依赖于区域文件中的名称服务器信息及 WHOIS 数据库的可用性。

通过使用黑名单进行钓鱼检测，可以准确地识别已被确认的网络钓鱼，大大降低了误检率，另一方面，黑名单还具有主机资源需求低的优点[61]。但是，由于大多数网络钓鱼活动的存活周期短，黑名单的方法在防御 0-hour 钓鱼攻击（新出现的钓鱼攻击）方面的有效性并不高。Sheng 等[62]的研究显示，黑名单的方法仅能检测 20%的 0-hour 钓鱼攻击，主要有以下 2 个原因。

1) 黑名单的加入过程造成延迟。一个新钓鱼活动的 URL、IP 地址等信息必须在确认其为网络钓鱼后才能加入黑名单，而像 PhishTank、MalwarePatrol 多提供黑名单的机构往往采用人工投票确认的方式判定一个可疑的活动是否是网络钓鱼，因此带来一定的延时。研究表明，大约 47%～83%的网络钓鱼在被发现 12 h 之后才能加入黑名单，但事实上，63%的网络钓鱼行为会在发生后的 2 h 内结束[62]。这一延迟极大地影响了黑名单方法检测的准确率。

2) 黑名单的更新造成延迟。黑名单的更新有 2 种方法：① 将更新的黑名单列表推送到客户端；② 服务器检查所访问的 URL 是否是钓鱼网站，然后将结果通知给客户端[63]。这 2 种方法都存在一定的问题。如果黑名单服务器广播更新的网络钓鱼黑名单，广播的频率低会产生延迟问题，频率过高又会增加服务器的负载。而第 2 种方法需要每个客户端联系黑名单服务器获取结果，虽然没有延迟问题，但可能会面临服务器的可扩展性问题。

3.2 启发式钓鱼检测

网络钓鱼的启发式检测是根据网络钓鱼之间的相似性，从已检测到的网络钓鱼攻击中提取一个或多个特征。虽然并不能保证在钓鱼攻击中总是存在这些特征，但是一旦识别出一组泛化的启发式特征，就可以实现 0-hour 钓鱼攻击检测，这是黑名单的方法所不具有的优点。但是，这种检测方式可能会增加将合法的网页或邮件误检的风险。

大多数启发式钓鱼检测使用的特征是从URL和HTML DOM（文档对象模型）中提取的[28]。Zhang 等[44]提出的基于内容的方法 CANTINA 是著名的基于启发式的检测方法之一。该方法通过计算网页页面内容的 TF-IDF 得到页面的词汇签名（排名最高的 5 个关键词），使用 Google 搜索引擎检索这 5 个关键词及当前域名（如 http://www.ebay.com/xxxx，则当前域名为“eBay”），根据检索返回的结果（若返回 0 条结果，则认为该行为是钓鱼）以及其他的启发式特征（表 2）判断页面是否合法。在该方法中，启发式规则的使用在一定程度上降低了误检率，但增加了漏检率。

表 2 CANTINA 使用的启发式规则

启发式规则	可疑的钓鱼
域名的注册时间	≤12 个月
已知图片	页面包含已知的 logo 但其域名不属于 logo 的所有者
可疑的 URL	URL 包含@或-
可疑的链接	页面中的链接包含@或-
IP 地址	URL 中包含 IP 地址
URL 中的“.”	URL 中“.”的个数≥5
表单	页面包含文本输入字段
TF-IDF-Final	将 TF-IDF 值最高的前 5 个词语和当前页面的域名作为页面的词汇签名在 Google 中检索，检查前 30 条检索结果中是否有与当前域名匹配的，若无检索结果，则认为是钓鱼

Lin 等[64]基于主流合法网站往往提供 2 个版本（移动版本和桌面版本）的网站服务，而网络钓鱼网站通常没有这一发现，针对多数网站单独构建移动端网站的情况，提出了基于用户设备检测的方法。该方法采用新的启发式规则，通过使用不同的用户代理（user agent）字符串对 URL 进行访问，比较返回的结果。若相同，说明该站点没有检测用户设备的机制，即该网站只有一个版本。若不同，则说明该站点有检测用户设备的机制。该方法虽然召回率较高（99%），但无法准确识别自适应网页设计（RWD, respond Web design）构建的合法网站，因此存在较高的误检率（15%）。

与黑名单的方法相比，基于启发式的检测方法能够检测新出现的网络钓鱼活动，但其误检率普遍高于黑名单[62]。这种方法比较简单，常以插件的形式应用于各种主流浏览器（如 Chrome、火狐、IE 浏览器等）上。然而，由于启发式的规则

特征主要来自于网络钓鱼的统计特征或人工总结，该类方法一方面依赖于领域知识，规则更新困难；另一方面，许多合法内容（如合法邮件、合法网页等）也有可能具有规则中的某些特征，从而造成误检率的提高。

3.3 基于视觉相似性的钓鱼检测

与其他方法不同，基于视觉相似性的钓鱼检测并不关注底层的代码或网络层面的特征，而是通过比较页面之间视觉特征（局部特征和全局特征）来实现网络钓鱼检测。通常这种方法包括 2 个部分：视觉特征提取和相似性度量。从待检测网页提取一组特征，然后基于该特征集，计算该网页与数据库中所有网页之间的相似度得分。如果相似度得分超过某一阈值且该网页与合法网页信息数据库中的信息（域名等）不一致，则认为其是钓鱼网页。

基于视觉相似性的钓鱼检测分为基于 HTML 文本的匹配[37,38,40]和基于图像的匹配[41,42]。2005 年，Liu 等[37,38]提出了通过比较钓鱼网站和非钓鱼网站的视觉相似度进行网站类型判断的方法。该方法利用 HTML DOM 树，根据“视觉提示”将网页页面分块，然后使用 3 个度量评估待检测网站和合法网站之间的视觉相似性：块级相似性、布局相似性和风格相似性。如果一个网页的任何一个度量的值超过了预先设定的阈值，则该网页被认为是钓鱼网页。该方法能够以很低的误检率完成网络钓鱼的检测，虽然在进行页面之间的相似度计算时速度很快，但在合法页面视觉信息数据库数据量很大时，对页面进行判定的耗时会很严重。而且该方法很大程度上取决于网页分割的结果，尤其是块级相似性和布局相似性的计算，因此该方法的检测效果依赖于 DOM 表示的可用性，无法检测具有相似的外观、但 DOM 表示不同的网页。

在 2006 年，Fu 等[41]提出了一种使用陆地移动距离（EMD，earth mover’s distance）衡量网页页面视觉相似度的方法。该方法首次将网页页面映射为低分辨率的图像，然后使用颜色特征和坐标特征表示图像的特征。利用 EMD 计算网页页面图像之间的特征距离，并训练一个 EMD 阈值向量对页面进行分类。该方法完全基于 Web 页面的图像特征，不依赖于 HTML 内容的可用性。但是由于可疑网页和合法网页的数量巨大，一些不相关的网页图像对也可能具有高相似度，导致误检率的增加。

但 Fu 等的方法仅考虑网页图像中的颜色及其分布特点，未考虑网页中不同部分之间的位置关系，这可能导致相似检测的失效。针对该问题，曹玖新等[42]提出了基于嵌套 EMD 的钓鱼网页检测算法，对图像进行分割，抽取子图特征并构建网页的特征关系图（attributed relational graph），计算不同 ARG 属性距离并在此基础上采用嵌套 EMD 方法计算网页的相似度。

现有的基于视觉相似性的钓鱼检测很大程度上依赖于网站快照的白名单或黑名单的使用[61]。从理论上讲，该方法是一种泛化的黑名单或白名单，需要频繁更新以保持完整性。另一方面，该方法往往假设钓鱼网站与合法网站相似，但在实际应用中，这种假设并不总是成立。对于只是部分复制合法网站（小于 50%）的钓鱼网站，基于视觉相似性的方法将无法成功检测[65]。

3.4 基于机器学习的钓鱼检测

机器学习是人工智能的一个分支，基于机器学习的钓鱼检测将网络钓鱼检测问题视为一个文本分类或聚类问题，然后运用各种机器学习中的分类算法（如 *K*-近邻、C4.5、支持向量机、随机森林等）、聚类算法（如 *K*-means、DBSCAN 等）达到对网络钓鱼攻击进行检测和防御的目的。目前，机器学习方法主要分为有监督学习、半监督学习和无监督学习 3 种，因此基于机器学习的钓鱼检测也是使用这 3 类学习方法实现的。

3.4.1 有监督学习方法

基于有监督学习方法的网络钓鱼检测是利用带标记的钓鱼数据（钓鱼邮件、钓鱼网站、钓鱼 URL 等）和带标记的合法数据训练得到一个分类器，通过得到的分类器对待检测数据进行分类的方法，其整体流程如图 8 所示。

在网络钓鱼检测中常用的有监督学习方法有随机森林（random forest）、序列最小优化算法（SMO, sequential minimal optimization）、J48、朴素贝叶斯等，其简要介绍如下。

随机森林：由多个决策树分类器组成，每棵树的特征是总特征集合中随机的一组、样本数据是整体样本数据有放回采样的集合，该算法最终的判决结果由所有个体决策树投票决定[66]。

SMO：由 John Platt 设计的用于训练支持向

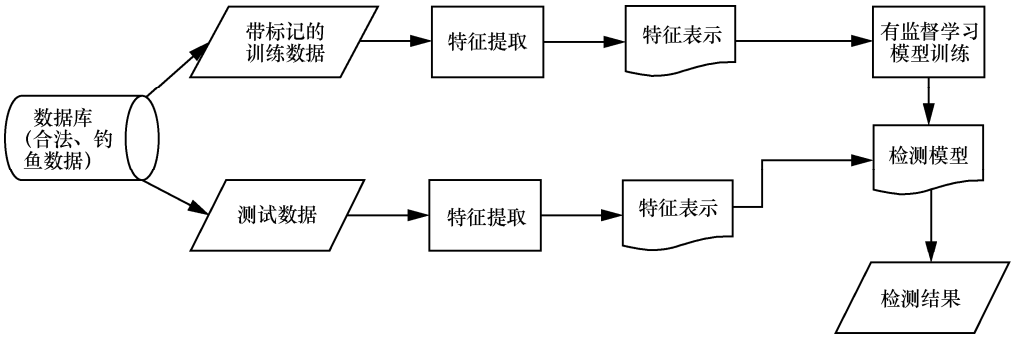


图 8 基于监督学习的钓鱼检测流程

量分类器的序列最小优化算法[67]。

J48 算法：是 C4.5 分类算法的 Java 实现[68]。

朴素贝叶斯：是一个应用贝叶斯定理的简单分类器，该方法严格假定条件独立[69]。

在文献[17, 19, 20]中分析比较了上述 4 种算法用于网络钓鱼检测的效果，结果表明，在提取的特征相同的情况下，J48 和随机森林这 2 个算法的效果普遍较好。但随机森林在合法实例和钓鱼实例权重变化时，加权错误率波动较大[70]。

对于网络钓鱼的检测来说，分类的准确性主要取决于在分类的学习阶段所识别的网络钓鱼特征[18]。因此，在大多数使用机器学习技术进行钓鱼检测的研究中，其关注的重点大多是如何选择更有效的特征才能训练出准确率高、具有顽健性、能处理 0-day 钓鱼攻击的分类器。

Xiang 等在 CANTINA[44]的基础上提出了 CANTINA+的检测方法[71]，该方法主要分为 3 个阶段：首先，利用 HTML DOM、搜索引擎及第三方服务提取了揭示网络钓鱼攻击特点的 8 个新颖的特征；然后，在进行分类过程之前，使用启发式规则过滤掉没有登录框的网页；最后，使用机器学习算法对 URL 词汇特征、Form 表单、WHOIS 信息、PageRank 值搜索引擎检索信息等 15 个具有高度表达性的钓鱼特征进行学习，实现钓鱼网页的分类。

Marchal 等指出[27]：1) 尽管钓鱼者试图使钓鱼页面与目标页面尽可能地相似，但是他们在搭建钓鱼页面时存在一定的约束；2) 网页可以由来自网页不同部分的一组关键词（如正文文本、标题、域名以及 URL 的一些内容等）表征，但合法网页和钓鱼网页使用这些关键词的方式是不同的。基于这 2 个观点，他们提出了一种用于检测钓鱼网站和目标的新方法，选取了 212 个特征（如表 3 所示），然后使用 Gradient Boosting 进行钓鱼网站的检测。该方法不需要大量训练数据就可以很好地扩展到更大的测试数据，具有不依赖于语言、品牌，速度快，可以自适应钓鱼攻击及可完全在客户端实现的优点。但是该方法对基于 IP 的钓鱼 URL 进行检测时精度太低，并且可能将空的或不可用的网页以及保留域名误分为钓鱼。

表 3 特征集

标号	数目	类型
f_1	106	URL
f_2	66	词语使用的一致性
f_3	22	启动和登录主级域（mld）的使用
f_4	13	注册域名（RND）的使用
f_5	5	网页页面内容
f_{all}	212	整个特征集

Moghimi 等[72]则是在有监督学习的基础上，提出了一种基于规则的网上银行钓鱼攻击检测的方法，该方法首先使用支持向量机算法（SVM，support vector madisone）训练网络钓鱼的检测模型，随后使用 SVM_DT 算法提取隐藏的决策规则，构建决策树。该方法仅用 10 条规则就达到了很高的精度和敏感性（准确率：98.86%，*F*1：0.989 98，灵敏度：1）。同样，该方法也存在缺点，它完全依赖页面内容，并且假设钓鱼网站的页面只使用合法页面的内容，因此难以检测识别钓鱼攻击者重新设计的钓鱼网站。

3.4.2 半监督学习方法

有监督学习方法（如 SVM、朴素贝叶斯等）通常需要大量的数据进行模型的训练，才能达到

很高的准确率。在网络钓鱼的标记样本很少时，无法使用监督学习的方法，在这种情况下往往采用半监督学习（如图 9 所示）或无监督学习的方法。

2016 年，Han 等[8]针对鱼叉式网络钓鱼活动（spear phishing）的标记数据数量有限这一问题，提出了基于邮件 profiling 特征的鱼叉式网络钓鱼活动的归因和识别模型。他们选取了邮件的四类 profiling 特征：来源特征、文本特征、附件特征和收件人特征，这些特征不仅能充分反映鱼叉式网络钓鱼邮件特征，而且对钓鱼邮件活动的演变具有顽健性。在此基础上，Han 等提出了基于属性图的半监督学习（SSL，semi-supervised learning）框架，提高了机器学习算法在标记邮件有限的情况下进行鱼叉钓鱼活动归因和识别的实用性。

图 10 是钓鱼活动归因模型的整体工作流程[8]，流程图中的每一个分析模块都执行相同的半监督学习过程。他们根据邮件的 profiling 特征构造 *K*-近邻属性图。在属性图中，每个节点代表一封邮件，节点之间的边代表两者的相似性。系统在属性图中传递标签信息，并将邮件归因于相应的活动。实验表明，该模型在已知活动的归因中，仅使用 25 封标记邮件，就达到了 0.9 的 *F*1 值、0.01 的误检率；同时，该模型还可以检测未知的鱼叉式网络钓鱼，在实验中使用 246 封标记邮件检测到了 100%的 darkmoon 活动、超过 97%的 samkams 活动以及 91%的 bisrala 活动。

与监督学习方法相比，半监督学习方法仅需要少量的训练样本，能充分利用大量的未标记样本实现网络钓鱼的检测和识别，减少了人工标记数据的工作量。但是基于半监督学习的检测往往会比基于有监督学习的检测准确率低，特别是在未标记样本的分布与有标记样本的分布差异较大的情况下，钓鱼检测的性能会受到很大影响。

3.4.3　无监督学习方法

图 11 为基于无监督学习的钓鱼检测的流程。在无监督学习中，事先不需要任何训练样本，即不需要标记数据，直接对数据进行建模。*k*-means

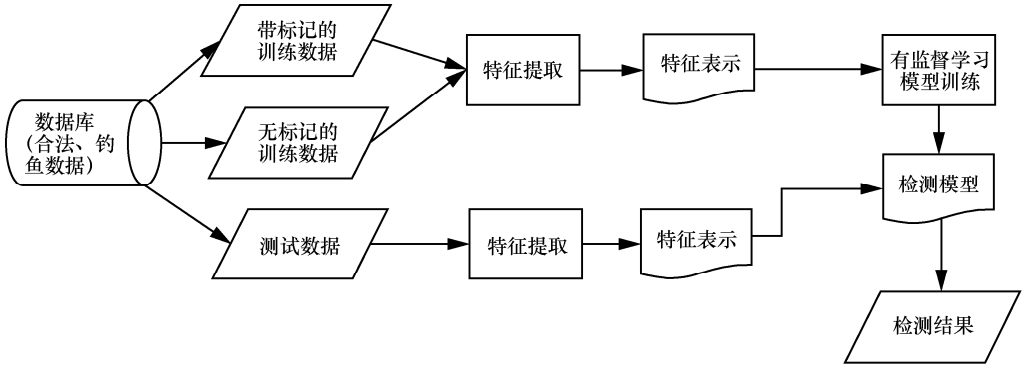


图 9　基于半监督学习的钓鱼检测流程

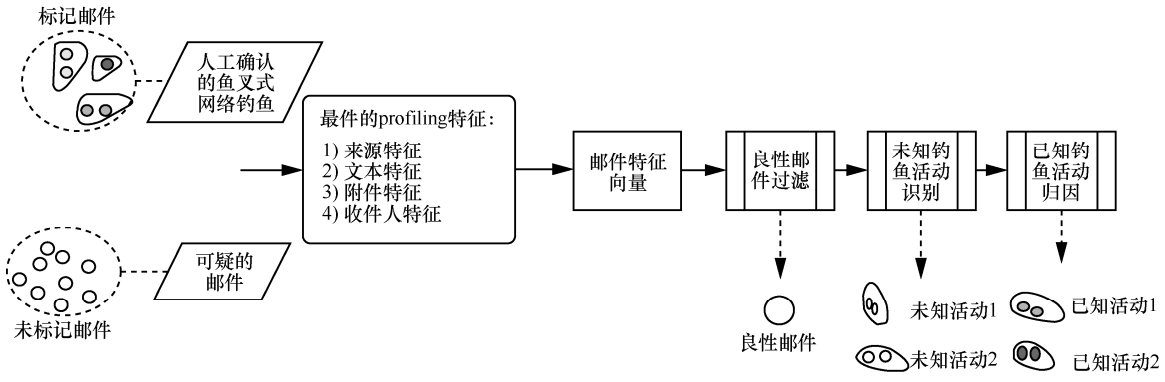


图 10　鱼叉式钓鱼邮件分析流程

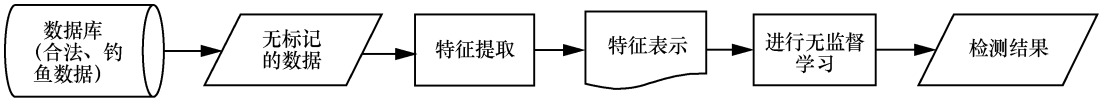


图 11 基于无监督学习的钓鱼检测流程

和 DBSCAN（density-based spatial clustering of application with noise）是常用的无监督学习算法。*k*-means 算法通过随机设置 *k* 个聚类中心来构建 *k* 个簇，然后将实例迭代地划分到距离（如欧氏距离）最近的聚类中心所在的簇并更新聚类中心。重复该迭代过程直至收敛。

DBSCAN 基于实例的密度划分实例，与 *k*-means 不同的是，它不需要事先确定簇的数量。2010 年，Liu 等[73]以网页页面之间的链接关系、检索结果的排序关系、文本相似性及页面布局相似性等关系作为特征，采用 DBSCAN 聚类算法对钓鱼网页进行识别。基于无监督学习的网络钓鱼检测减少了人工标记的代价，但检测的准确率不高且检测结果受数据集的结构影响较大。

3.5 基于自然语言处理技术的钓鱼检测

自然语言理解是计算机科学的一个领域，它使计算机能够理解人类所讲的语言，也就是说，让计算机以一种有意义的方式处理自然语言中的数据和指令。Verma 等[74]于 2012 年提出利用自然语言处理技术解决网络钓鱼邮件检测问题。

Aggarwal 等[17]针对电子邮件沟通方式的钓鱼活动，提出了检测不包含任何链接的网络钓鱼邮件的方案，这些邮件往往是利用用户的好奇心，促使用户向钓鱼者回复敏感信息。该检测方法使用自然语言处理和 WordNet [注9]实现。通过对钓鱼邮件的分析，Aggarwal 等提取了不包含链接的网络钓鱼邮件所共有的要素：缺少收件人的名字、提及钱、诱导回复的句子以及紧迫感。通过对邮件文本进行词性分析和词干提取，得到以下打分标准。

$$Score(r)=\frac{n(m+s+u)}{2^{L}} \quad \forall r\in SR \tag{11}$$

其中，

R 是一个表示要求回复邮件的词的集合。

SR 表示 *R* 的同义词集合中的词的后续 4 个下义关系词的同义词集合。

若邮件中没有提到收件人的姓名，*n*=1，否则 *n*=0。

若邮件中提到钱，*m*=1，否则 *m*=0。

若邮件中有 *SR* 中的词，*s*=1，否则 *s*=0。

若邮件中有 *SR* 中的词的句子同时有一种紧迫的语气，*u*=1，否则 *u*=0。

L=从 *R* 中的词到达词 *r* 的下义链接的数目。

每封邮件的最终得分= $\max_{r\in\text{邮件 and } r\in SR} Score(r)$ 。

该方法可以很好地检测电子邮件沟通式的钓鱼邮件，但无法处理电子邮件中包含的附件。对于包含附件的电子邮件，可以将其他技术（如光学字符识别技术）与该方法相结合，提取附件和邮件文本内容特征进行钓鱼邮件的检测。

此后，Yasin 等在文献[19]中提出了钓鱼相加权的概念，使用知识发现与机器学习分类算法相结合的方法进行网络钓鱼邮件的检测。从整体上来说，它与大多数基于机器学习的钓鱼检测方法的流程是一致的，首先基于语料库进行特征选择、特征提取，然后基于提取的特征训练模型，再将训练得到的模型用于分类决策。不同之处在于特征选择的过程（即预处理阶段），这个阶段通过以下 4 个步骤完成对邮件标题、邮件正文以及文本特征的提取：1) 文本解析、标记和词干提取；2)去除停用词；3) 语义文本处理；4) 钓鱼项加权。

在语义文本处理的过程中，根据同义词和词义的上下义关系，电子邮件中的每个词块都使用其与 WordNet 本体中概念相关的词语进行了扩展。这个过程有助于识别不同的电子邮

注9 WordNet 是一个包含语义信息的英文字典，它根据词条的意义将它们分组，每一个具有相同意义的词条组称为一个 synset（同义词集合）。WordNet 为每一个 synset 提供了简短、概要的定义，并记录不同 synset 之间的语义关系。http://wordnet.princeton.edu。

件消息中的标记之间的语义关系，缩短彼此接近的特征向量之间的距离，进而提高分类精度。

与其他方法相比，基于自然语言处理技术（natural language process）的检测方法在网络钓鱼检测的研究中并不常见，这可能与缺少比较成熟的自然语言处理技术有关。另一方面，很多电子邮件的内容可能包含打字错误，使用 NLP 处理起来更为复杂。

4　网络钓鱼检测方法对比分析及面临的挑战

4.1　网络钓鱼检测方法对比分析

任何一种单一的技术都无法满足钓鱼检测的所有需求。本节选择了代表性的反钓鱼工作进行对比分析，从所属类别、基本原理及优缺点等方面进行了分析和总结，便于更直观地说明各类钓鱼检测工作的特点，并为后续研究提供明晰的参考（如表 4 所示）。

表 4　　**网络钓鱼检测技术比较**

类别	典型工作	优点	缺点	基本原理
黑名单	PhishNet[59]	• 扩充了黑名单列表，可检测部分未出现在黑名单中的钓鱼 URL • 设计简单，易于实现	• 无法检测 0-hour 钓鱼攻击 • 严重依赖原始的黑名单 • 借助第三方的工具进行 DNS 查询和页面内容匹配，可能引起较大的带宽开销和时间开销	基于原始黑名单生成新的 URL，扩充了黑名单列表，并通过 URL 分解和相似性计算进行钓鱼 URL 的检测识别
	基于域名黑名单的方法[60]	• 缩短了域名加入黑名单的时间	• 依赖于区域文件的可用性 • 对 WHOIS 数据库的访问可能会成为整个方法的瓶颈	基于恶意域名及其 NS 往往是新的、通常一起管理的启发式想法，利用 zone file 的 NS 信息和 WHOIS 信息实现对域名黑名单的主动扩充
启发式	CANTINA[44]	• 可检测 0-hour 钓鱼攻击 • 容易实现	• 具有语言依赖性，TF-IDF 对东亚语言的处理效果不好 • 查询 Google 会带来时间开销，影响性能 • 规则简单，易规避	通过使用 TF-IDF 算法及 Google 检索结果，结合其他启发式规则（域名注册时间、URL 中点的个数等）实现对钓鱼 URL 的检测识别（使用了 URL 和 HTML DOM 特征)
	基于用户设备检测的方法[64]	• 可检测 0-hour 钓鱼攻击 • 方法简单，易实现	• 无法准确识别自适应网页设计（RWD）构建的合法网站 • 规则过于简单，很容易规避	提出了新的启发式规则，主流合法网站往往具有移动和桌面 2 个版本的网站服务，而钓鱼网站通常没有，并基于此规则，结合 SVM 算法进行网络钓鱼的检测和识别（使用了 HTML 特征）
视觉相似性	Liu 等[37]	• 可检测 0-hour 钓鱼攻击 • 对与合法页面视觉及 DOM 表示相似的网络钓鱼检测效果很好	• 依赖于 DOM 的可用性 • 使用的合法页面的视觉信息列表的完整性和时效性对钓鱼检测的结果影响较大 • 使用图片特征，效率较低	使用块级相似性、布局相似性和风格相似性 3 个度量来衡量待检测页面与合法页面之间的视觉相似性，从而判别该页面是否是网络钓鱼页面（使用了网页页面特征）
	基于 EMD 的视觉相似度方法[41]	• 可检测 0-hour 钓鱼攻击 • 对具有视觉相似性的钓鱼，检测准确率高 • 不依赖于 HTML 的可用性	• 无法检测与目标网页视觉上不相似的钓鱼网站 • 需要存储计算大量的合法页面的图像信息	将网页图像映射为低分辨率的图像，使用颜色和坐标对图像进行特征表示，利用陆地移动距离计算网页图像之间的特征距离，根据 EMD 值完成网络钓鱼的检测识别（使用了网页页面图像特征）
	基于嵌套 EMD 的钓鱼网页检测方法[42]	• 具有较好的顽健性 • 不依赖于 HTML 的可用性 • 考虑了页面中各部分的相对位置因素	• 图像分割处理部分复杂度较大	将网页图像分割，抽取子图特征并构建网页的 ARG，在计算不同 ARG 属性距离的基础上使用嵌套 EMD 算法计算网页相似度

续表

类别	典型工作	优点	缺点	基本原理
机器学习	CANTINA+[71]	• 可检测 0-hour 钓鱼攻击 • 在分类之前使用启发式规则进行过滤，提高了效率 • 对所使用的特征进行了性能分析	• 使用了 HTML DOM 和第三方服务，受其可用性的限制 • 使用了搜索引擎，可能会影响性能	利用 HTML DOM、搜索引擎和第三方服务提取了 8 个新特征，使用机器学习算法完成钓鱼检测，同时基于启发式规则实现了 2 个过滤器以降低误检率、提高运行速度（使用了 URL、HTML 特征）
	一种用于检测钓鱼网站和目标的新方法[27]	• 可检测 0-hour 钓鱼攻击 • 准确率、召回率、精度都很高 • 不需要大量数据 • 具有语言独立性 • 可完全在客户端实现	• 对于空的或不可用的网页和保留域名可能产生误判 • 对基于 IP 的钓鱼 URLs 的分类精度太低	基于钓鱼攻击者在搭建钓鱼页面时的约束及合法网页和钓鱼网页使用关键字的方式不同这两点，提取了 212 个特征，并使用 Gradient Boosting 进行钓鱼网站的检测（使用了 URL 和 HTML 特征）
	PhishDetector[72]	• 可检测 0-hour 钓鱼攻击 • 可从分类模型中提取隐含的知识，可与启发式方法结合 • 不依赖第三方的服务（搜索引擎、浏览器历史等）	• 完全依赖页面内容 • 无法检测使用 Flash 或者图片等（不使用 DOM）的钓鱼网页	使用 SVM 训练钓鱼检测模型，并使用 SVM_DT 算法提取分类精度很高的隐含规则。（使用了 URL、HTML DOM 特征）
	基于邮件 profiling 特征的鱼叉式网络钓鱼活动的归因与识别[8]	• 可检测 0-hour 钓鱼攻击 • 不需要大量的标记数据，降低人工标记开销 • 高检测率，低误检率	• 算法复杂，计算开销较高	提出了基于属性图的半监督学习框架，实现对鱼叉式网络钓鱼活动的归因和识别（使用了邮件特征）
	基于 DBSCAN 的方法[73]	• 可检测 0-hour 钓鱼攻击 • 不需要标记数据	• 使用了搜索引擎，可能会有时间开销或检索方面的问题	利用网页页面之间的链接关系、检索结果的排序关系、文本相似性及页面布局相似性等特征，采用 DBSCAN 聚类算法进行钓鱼检测。
自然语言处理	基于词性分析和词干提取的方法[17]	• 可有效识别不含链接网络钓鱼邮件	• 只针对邮件文本内容，无法检测附件内容 • 依赖于已知的钓鱼邮件，无法检测 0-hour 钓鱼攻击	针对不包含任何链接的网络钓鱼邮件，通过对邮件文本进行词性分析和词干提取，然后根据该类邮件所共有的特征对待检测邮件进行打分来判断其是否是钓鱼邮件。（使用了邮件特征）
	知识发现与机器学习结合的检测方法[19]	• 精度高 • 使用的特征较少	• 不具有自适应机制	提出了钓鱼相加权的概念，将自然语言处理中对文本处理的技术与机器学习结合起来进行网络钓鱼的检测和识别

在前文介绍的钓鱼检测评价指标中，最重要的 2 个是网络钓鱼攻击的检测精度和误检率。绝大多数的网络钓鱼攻击的存活时间都很短，因此提高对新出现的钓鱼攻击的检测能力十分必要的。而一个网络钓鱼检测系统的误检率的高低则直接关系到用户对该系统的信赖程度。

基于黑名单的钓鱼检测可以准确识别已被确认的网络钓鱼，查找效率高、快速精准，适用于要求误检率很低的情况。黑名单的方法设计简单易实现，但由于黑名单的加入和更新存在延迟，往往很难满足正确性、及时性和完整性这 3 个要求，容易产生漏检的情况，也无法检测新出现的网络钓鱼攻击。另外，黑名单的构建和更新需要人工干预和验证，可能消耗大量的资源。黑名单的方法虽然不适合单独使用，但是可以和其他能够检测 0-hour 钓鱼攻击的方法（如启发式的方法、基于视觉相似性的方法等）结合使用，在将误检率控制在可接受的范围内的同时，提高对新出现的钓鱼攻击的防御能力。

启发式钓鱼检测可在网络钓鱼攻击发起时就

进行，不必等待黑名单的更新，因此可以实现 0-hour 网络钓鱼攻击的检测识别。并且这类方法简单、易于实现，在一些主流浏览器（如 Chrome、火狐、IE 等）上得到广泛应用，但这种通过统计特征或人工总结得到的启发式规则有很大的局限性，一些合法网站也可能具有所使用的启发式规则的某些特征，导致误检率的增加。此外，启发式的规则简单，网络钓鱼攻击者可以通过重新设计钓鱼攻击，很容易规避启发式的钓鱼检测。

基于视觉相似性的钓鱼检测是基于钓鱼页面往往与合法页面在视觉上相似这一假设实现的，针对性强，可以很好地解决由图片构成的钓鱼网站的检测问题，也能够防御新出现的网络钓鱼攻击，但其本质上仍是黑名单的方法，需要频繁地更新，保持数据库的完整和最新，才能维持有效性。另一方面，这种使用图像特征的方法需要对图像信息进行处理，并且需要计算待检测页面与所有合法页面之间的视觉相似度，检测效率较低，与其他方法相比，需要更多的计算和存储成本。

基于自然语言技术的钓鱼检测通过让机器“理解”网络钓鱼邮件或钓鱼网站的内容，从语义的角度实现网络钓鱼的检测，但是目前相关研究较少，并且自然语言处理技术虽然对英文等拉丁语系的语言处理效果较好，但对中文语义的理解方面仍存在很大的问题，需要进一步发展完善。

将网络钓鱼问题抽象为一个分类或聚类的问题，然后采用机器学习算法完成分类或聚类任务，是目前网络钓鱼检测常用的手段之一。通过利用已有数据构建模型，减少了大量的人力，提高了钓鱼检测的效率。基于机器学习的检测方法还可实现 0-hour 网络钓鱼攻击检测。另外，机器学习的方法可以从各个维度的特征（如 URL 特征、HTML 特征、视觉特征等）进行学习，并方便基于新的钓鱼形式进行特征空间的拓展，提高了检测精度；具有可扩充性，可通过增量学习将新的钓鱼数据加入数据集对检测模型进行修正；强化学习等技术可以不断提高分类器的能力，从而达到自适应网络钓鱼攻击发展的目的。

4.2　网络钓鱼检测面临的挑战

尽管研究者们已经研究开发了诸多网络钓鱼检测技术、工具来帮助用户检测和避免网络钓鱼，然而网络钓鱼的攻击和防御之间的博弈从未停止。互联网的迅速发展也给网络钓鱼检测带来了很大的挑战。

1) 网页规模迅速由 GB 级、TB 级向 PB、ZB 级扩大，对网络钓鱼检测技术的存储、计算能力的要求增大。

2) 攻击者搭建钓鱼网页成本降低，给攻击者持续缩短网络钓鱼活动的生命周期带来了便利。

3) 网络钓鱼不再局限在计算机层面，手机平台成为网络钓鱼的新目标。2012 年趋势科技（trend micro）的研究发现了 4 000 条为手机网页设计的钓鱼 URL[75]。尽管这个数字不到所有钓鱼 URL 的 1%，但它表明手机平台开始成为网络钓鱼攻击的新目标，并且由于手机屏幕的大小限制，手机网络钓鱼更具有欺骗性。

4) 传播途径不再局限于电子邮件、手机短信的方式，各种社交网站（如 Twitter[76]、微博）、网络游戏[77]、二维码[78]等的兴起使传播途径更多元化，也让网络钓鱼检测更困难。

5) 网络钓鱼攻击的形式繁多，鱼叉式网络钓鱼攻击、执行长欺诈、域欺骗（pharming）、标签钓鱼[79]（tabnabbing）等各种攻击形式层出不穷，难以应对。

6) DNSsec 协议推动较为缓慢，钓鱼攻击者常常利用名址解析存在的漏洞，劫持合法网站展开钓鱼活动。这种网站劫持的钓鱼攻击，在用户访问合法网站时跳转到钓鱼网站，用户往往难以察觉，为钓鱼检测增加了难度。

除了客观环境给网络钓鱼检测带来的挑战外，攻击者们还会不断地改进攻击手段以规避检测，例如，使用对短链接技术[80]模糊钓鱼 URL 以更好地传播钓鱼链接；对网页内容进行各种混淆、加密；使用 Fast flux 技术规避黑名单技术；采用人机识别技术对访问者的身份进行判定，只有在认定是人工浏览行为时才推送钓鱼网页，否则推送事先准备好的合法网页（如百度首页）；进一步缩短网络钓鱼行为的生命周期等[81]。

5　结束语

本文从定义、发展趋势、攻击目的等方面对

网络钓鱼进行了概述，并对常用的网络钓鱼检测方法进行了分析总结。虽然目前已经有很多效果不错的检测方法，但网络钓鱼的攻击与防御就是一场“军备竞赛”。随着检测技术的发展，攻击者们也不断地设计出新的钓鱼形式以规避已有的检测技术。正如“开发商只有在黑客找到他们之后才纠正他们的错误”，人们无法知道网络钓鱼攻击者下一个攻击的手段是怎样的，因此，如何使检测方法自适应网络钓鱼的发展演化是网络钓鱼检测方法研究的关键所在。

从目前的发展现状来看，机器学习存在很大的发展潜力。机器学习的方法具有对高维特征进行学习的能力，检测效果较好。而且这类方法具有很好的可扩充性，只需将新的钓鱼数据加入数据集就可完成对钓鱼检测模型的修正，因此能够很好地适应网络钓鱼攻击的发展，实现 0-hour 网络钓鱼攻击检测。但是，目前基于机器学习的网络钓鱼检测方法中往往缺乏对各个特征效果的有效评估，无法确定每个特征对钓鱼检测的贡献如何。盲目地使用高维度的特征，可能会出现付出了很高的计算代价，但检测效果却只有略微提升的情况。本文认为，这是机器学习的检测方法在之后的发展中所需要解决的问题。另一方面，基于视觉相似性的钓鱼检测可以很好地解决由图片构成的钓鱼网站的检测问题，这类方法大部分依赖于图像的相似性检测。近年来，深度学习日益火热，极大地促进了图像处理效果的提高。结合基于视觉相似性的钓鱼检测的思想，将深度学习技术应用于网络钓鱼检测也将成为今后的研究方向之一。此外，随着自然语言处理技术的发展成熟，基于此类技术的钓鱼检测方法也非常有前景。

参考文献：

[1] 国家网络空间安全战略[EB/OL]. http://news.xinhuanet.com/politics/2016-12/27/c1120196479.htm.
National cybersecurity strategy[EB/OL]. http://news.xinhuanet.com/politics/ 2016-12/27/c1120196479.htm.

[2] Anti-Phishing Working Group(APWG). Phishing activity trends report-second quarter 2016[EB/OL]. https://docs.apwg.org/reports/apwgtrendsreportq22016.pdf.

[3] WEIDER D Y, NARGUNDKAR S, TIRUTHANI N. A phishing vulnerability analysis of web based systems[C]//Computers and Communications. 2008: 326-331.

[4] E.M.C.Corporation.RSA monthly fraud report[EB/OL]. http://australia.emc.com/collateral/fraud-report/h13929-rsa-fraud-report-jan-2015.pdf.

[5] 中国反钓鱼网站联盟. 2012 年中国反钓鱼网站联盟年报[EB/OL]. http://apac.cn/gzdt/qwfb/201408/P020140826493067614020.pdf.
APAC. Coalition against phishing site report of China in 2012[EB/OL]. http://apac.cn/gzdt/qwfb/201408/P020140826493067614020.pdf.

[6] 中国反钓鱼网站联盟. 2016 年 9 月钓鱼网站处理简报[EB/OL]. http://apac.cn/gzdt/qwfb/201610/P020161110519501201415.pdf.
APAC. Phishing site processing presentation in september 2016[EB/OL]. http://apac.cn/gzdt/qwfb/201610/P020161110519501201415.pdf.

[7] 中国反钓鱼网站联盟. 2015 年 12 月钓鱼网站处理简报[EB/OL]. http://apac .cn/gzdt/qwfb/201601/P020160108491677785300.pdf.
APAC. Phishing site processing presentation in december 2015 [EB/OL]. http://apac.cn/gzdt/qwfb/201601/P020160108491677785300.pdf.

[8] HAN Y F, SHEN Y. Accurate spear phishing campaign attribution and early detection[C]//The 31st Annual ACM Symposium on Applied Computing. 2016: 2079-2086.

[9] ALARM S, EL-KHATIB K. Phishing susceptibility detection through social media analytics[C]//The 9th International Conference on Security of Information and Networks. 2016: 61-64.

[10] Krebs on security[EB/OL]. https://krebsonsecurity.com/2016/04/fbi-2-3- billion-lost-to-ceo-email-scams/.

[11] Anti-Phishing Working Group(APWG). Global phishing survey:trends and domainname use in 2H2014[EB/OL]. http://docs.apwg.org/reports/APWGGlobalPhishingReport2H2014.pdf.

[12] YAN G, EIDENBENZ S, GALLI E. Sms-watchdog: profiling social behaviors of SMS users for anomaly detection[C]//The International Workshop on Recent Advances in Intrusion Detection. 2009: 202-223.

[13] NASSAR M, NICCOLINI S, EWALD T. Holistic VoIP intrusion detection and prevention system[C]//The 1st International Conference on Principles, Systems and Applications of IP Telecommunications. 2007: 1-9.

[14] SONG J, KIM H, GKELIAS A. iVisher: real-time detection of caller ID spoofing[J]. ETRI Journal, 2014, 36(5): 865-875.

[15] 彭富明,张卫丰,彭寅. 基于文本特征分析的钓鱼邮件检测[J]. 南京邮电大学学报(自然科学版),2012(5): 140-145.
PENG F M, ZHANG W F, PENG Y. Detection of phishing emails based on text characteristic analysis[J]. Journal of Nanjing University of Posts and Telecommunication, 2012(5):140-145.

[16] HUSÁK M, CEGAN J. PhiGARo: automatic phishing detection and incident response framework[C]//Availability, Reliability and Security (ARES). 2014: 295-302.

[17] AGGARWAL S, KUMAR V, SUDARSAN S D. Identification and detection of phishing emails using natural language processing techniques[C]//The 7th International Conference on Security of Information and Networks. 2014: 217.

[18] AKINYELU A A, ADEWUMI A O. Classification of phishing email using random forest machine learning technique[J]. Journal of Applied Mathematics, 2014.

[19] YASIN A, ABUHASAN A. An intelligent classification model for phishing email detection[J]. 2016, 8(4):55-72.

[20] VERMA R, RAI N. Phish-IDetector: Message-ID based automatic phishing detection[C]//e-Business and Telecommunications (ICETE). 2015(4): 427-434.

[21] 黄华军, 钱亮, 王耀钧. 基于异常特征的钓鱼网站 URL 检测技术[J]. 信息网络安全, 2012,(01): 23-25,67.
HUANG H J, QIAN L, WANG Y J. URL Detecting technology of phshing site based on anomalous characte[J].Netinfo Security, 2012,(1): 23-25.

[22] BLUM A, WARDMAN B, SOLORIO T, et al. Lexical feature based phishing URL detection using online learning[C]//The 3rd ACM Workshop on Artificial Intelligence and Security. 2010: 54-60.

[23] MA J, SAUL L K, SAVAGE S, et al. Identifying suspicious URLs: an application of large-scale online learning[C]//The 26th Annual International Conference on Machine Learning. 2009: 681-688.

[24] MA J, SAUL L K, SAVAGE S, et al. Beyond blacklists: learning to detect malicious Web sites from suspicious URLs[C]//The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009: 1245-1254.

[25] FEROZ M N, MENGEL S. Examination of data, rule generation and detection of phishing URLs using online logistic regression[C]//2014 IEEE International Conference on Big Data. 2014: 241-250.

[26] FEROZ M N, MENGEL S. Phishing URL detection using URL ranking[C]//The IEEE International Congress on Big Data. 2015: 635-638.

[27] MARCHAL S, SAARI K, SINGH N, et al. Know your phish: Novel techniques for detecting phishing sites and their targets[C]// Distributed Computing Systems (ICDCS). 2016: 323-333.

[28] RAMESH G, KRISHNAMURTHI I, KUMAR K S S. An efficacious method for detecting phishing webpages through target domain identification[J]. Decision Support Systems, 2014, 61: 12-22.

[29] ABRAHAM D, RAJ N S. Approximate string matching algorithm for phishing detection[C]//Advances in Computing, Communications and Informatics. 2014: 2285-2290.

[30] 何高辉, 邹福泰, 谭大礼, 等. 基于 SVM 主动学习算法的网络钓鱼检测系统[J]. 计算机工程,2011,(19):126-128.
HE G H,ZOU F T,TAN D L, et al. Phishing detection system based on SVM active learning algorithm[J]. Computer Engineering, 2011(19): 126-128.

[31] CHIEW K L, CHANG E H, TIONG W K. Utilisation of website logo for phishing detection[J]. Computers & Security, 2015, 54: 16-26.

[32] GENG G G, LEE X D, ZHANG Y M. Combating phishing attacks via brand identity and authorization features[J]. Security and Communication Networks, 2015, 8(6): 888-898.

[33] 胡向东,刘可,张峰,等. 基于页面敏感特征的金融类钓鱼网页检测方法[J]. 网络与信息安全学报, 2016,2(2): 31-38.
HU X D, LIU K, ZHANG F, et al. Methods of Financial fishing Web test based on page sensitive characteristics[J]. Chinese Journal of Network and Information Security, 2016, 2(2): 35-42.

[34] GENG G G, LEE X D, WANG W, et al. Favicon-a clue to phishing sites detection[C]//eCrime Researchers Summit (eCRS). 2013: 1-10.

[35] PAN Y, DING X. Anomaly based web phishing page detection[C]// Computer Security Applications Conference. 2006: 381-392.

[36] ALKHOZAE M G, BATARFI O A. Phishing websites detection based on phishing characteristics in the webpage source code[J]. International Journal of Information and Communication Technology Research, 2011, 1(6).

[37] WENYIN L, HUANG G, XIAOYUE L, et al. Detection of phishing webpages based on visual similarity[C]//Special Interest Tracks and Posters of the 14th International Conference on World Wide Web. 2005: 1060-1061.

[38] WENYIN L, HUANG G, XIAOYUE L, et al. Phishing Web page detection[C]//Document Analysis and Recognition.2005: 560-564.

[39] 张卫丰, 周毓明, 许蕾, 等. 基于匈牙利匹配算法的钓鱼网页检测方法[J]. 计算机学报, 2010,(10): 1963-1975.
ZHANG W F, ZHOU Y M, XU L, et al. Financial fishing Web test based on Hungarian matching algorithm[J]. Chinese Journal of Computers, 2010(10): 1963-1975.

[40] 邹学强, 张鹏, 黄彩云, 等. 基于页面布局相似性的钓鱼网页发现方法[J]. 通信学报,2016(S1):116-124.
ZOU X Q, ZHANG P,HUANG C Y, et al. Detecting methods of phshing Web based on the page layout[J].Journal on Comunications,2016(S1):116-124.

[41] FU A Y, WENYIN L, DENG X. Detecting phishing Web pages with visual similarity assessment based on earth mover's distance (EMD)[J]. IEEE transactions on dependable and secure computing, 2006, 3(4).

[42] 曹玖新, 毛波, 罗军舟, 等. 基于嵌套 EMD 的钓鱼网页检测算法[J]. 计算机学报, 2009, (5): 922-929.
CAO J X, MAO B, LUO J Z, et al. Financial fishing Web test based on nesting EMD[J]. Journal of Computers, 2009 (5): 922-929.

[43] TAN C L, CHIEW K L. Phishing website detection using URL-assisted brand name weighting system[C]//Intelligent Signal Processing and Communication Systems (ISPACS).2014: 54-59.

[44] ZHANG Y, HONG J I, CRANOR L F. Cantina: a content-based approach to detecting phishing web sites[C]//The 16th International Conference on World Wide Web.2007:639–648.

[45] YAN Z, LIU S, WANG T, et al. A genetic algorithm based model for chinese phishing e-commerce websites detection[C]//The International Conference on HCI in Business, Government and Organizations. 2016: 270-279.

[46] 赵加林. 基于 K-Means 和 SVM 的流行中文钓鱼网站识别研究[J]. 软件导刊, 2016(4):176-178.
ZHAO J L. Study of popular Chinese phshing site identification based on K-Means and SVM[J].Software Guide, 2016(4): 176-178.

[47] ZHANG W, JIANG Q, CHEN L, et al. Two-stage ELM for phishing Web pages detection using hybrid features[J]. World Wide Web, 2016: 1-17.

[48] 徐欢潇, 徐慧, 雷丽婷. 多特征分类识别算法融合的网络钓鱼识别技术[J]. 计算机应用研究, 2017(4) :1129-1132.
XU H X, XU H, LEI L T. Phishing identification technology with multiple feature classification recognition algorithm[J].Application Research of Computers,2017(4):1129-1132.

[49] PhishTank[EB/OL]. http://www.phishtank.com/.

[50] Millersmiles[EB/OL]. http://www.millersmiles.co.uk/.

[51] Spamassassin public corpus[EB/OL]. http://spamassassin.apache.org/publiccorpus/.

[52] MalwarePatrol[EB/OL]. http://www.malwarepatrol.com/.

[53] Open directory[EB/OL]. http://www.dmoz.org/.

[54] Open directory project[EB/OL]. https://zh.wikipedia.org/wiki/.

[55] ABU-NIMEH S, NAPPA D, WANG X, et al. A comparison of machine learning techniques for phishing detection[C]//The anti-phishing working groups 2nd annual eCrime researchers summit.2007: 60-69.

[56] Spamhaus[EB/OL]. https://www.spamhaus.org/.

[57] SURBL[EB/OL]. http://www.surbl.org/lists.

[58] Google safe browsing api[EB/OL]. https://www.google.com/transparencyreport/safebrowsing/.

[59] PRAKASH P, KUMAR M, KOMPELLA R R, et al. Phishnet: predictive blacklisting to detect phishing attacks[C]//INFOCOM. 2010: 1-5.

[60] FELEGYHAZI M, KREIBICH C, PAXSON V. On the potential of proactive domain blacklisting[J]. LEET, 2010, 10: 6.

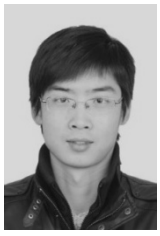
[61] KHONJI M, IRAQI Y, JONES A. Phishing detection: a literature survey[J]. IEEE Communications Surveys & Tutorials, 2013, 15(4): 2091-2121.

[62] SHENG S, WARDMAN B, WARNER G, et al. An empirical analysis of phishing blacklists[C]//The 6th Conference on Email and An-

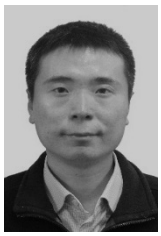
ti-Spam (CEAS). 2009.
[63] FLORÊNCIO D, HERLEY C. Analysis and improvement of anti-phishing schemes[C]//IFIP International Information Security Conference. 2006: 148-157.
[64] LIN I C, CHI Y L, CHUANG H C, et al. The novel features for phishing based on user device detection[J]. JCP, 2016, 11(2): 109-115.
[65] JAIN A K, GUPTA B B. Phishing detection: analysis of visual similarity based approaches[J]. Security and Communication Networks, 2017(4):1-20.
[66] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
[67] PLATT J C. 12 fast training of support vector machines using sequential minimal optimization[J]. Advances in Kernel Methods, 1999: 185-208.
[68] QUINLAN J R. C4. 5: programs for machine learning[M]. Elsevier, 2014.
[69] JOHN G H, LANGLEY P. Estimating continuous distributions in Bayesian classifiers[C]//The Eleventh Conference on Uncertainty in Artificial Intelligence.1995: 338-345.
[70] ABU-NIMEH S, NAPPA D, WANG X, et al. A comparison of machine learning techniques for phishing detection[C]//The anti-phishing Working Groups 2nd Annual eCrime Researchers Summit. 2007: 60-69.
[71] XIANG G, HONG J, ROSE C P, et al. Cantina+: A feature-rich machine learning framework for detecting phishing Web sites[J]. ACM Transactions on Information and System Security (TISSEC), 2011, 14(2): 21.
[72] MOGHIMI M, VARJANI A Y. New rule-based phishing detection method[J]. Expert Systems with Applications, 2016, 53: 231-242.
[73] LIU G, QIU B, WENYIN L. Automatic detection of phishing target from phishing webpage[C]//The 20th International Conference on Pattern Recognition (ICPR). 2010: 4153-4156.
[74] VERMA R, SHASHIDHAR N, HOSSAIN N. Detecting phishing emails the natural language way[C]//European Symposium on Research in Computer Security. 2012: 824-841.
[75] MICRO T. Mobile phishing: a problem on the horizon[EB/OL]. https://www.yumpu.com/en/document/view/10210640/rpt-monthly-mobile-review-201302-mobile-phishing-a-problem-on-the-horizon.
[76] JEONG S Y, KOH Y S, DOBBIE G. Phishing detection on Twitter streams[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2016: 141-153.
[77] ALBANESIUS C. Gaming apps increase spam, phishing by 50 percent[EB/OL]. http://www.pcmag.com/article2/0,2817,2362134,00.asp, 2010.
[78] VIDAS T, OWUSU E, WANG S, et al. QRishing: the susceptibility of smartphone users to QR code phishing attacks[C]//The International Conference on Financial Cryptography and Data Security. 2013: 52-69.
[79] SARIKA S, PAUL V. Parallel phishing attack recognition using software agents[J]. Journal of Intelligent & Fuzzy Systems, 2017, 32(5): 3273-3284.
[80] CHHABRA S, AGGARWAL A, BENEVENUTO F, et al. Phi.sh/$ ocial: the phishing landscape through short urls[C]//The 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference. 2011: 92-101.
[81] 沙泓州, 刘庆云, 柳厅文, 等. 恶意网页识别研究综述[J]. 计算机学报 2016(3):529-542.
SHA H Z,LIU Q Y,LIU T W,et al. Review of malicious Web recognition[J]. Journal of Computers,2016(3): 529-542.

作者简介：

张茜（1994-），女，河南杞县人，中国科学院大学硕士生，主要研究方向为网络应用与安全、下一代互联网技术。



延志伟（1985-），男，山西兴县人，博士，中国互联网络信息中心副研究员，主要研究方向为 IPv6 移动性管理、BGP 安全机制、信息中心网络架构。



李洪涛（1977-），男，河北保定人，中国互联网络信息中心高级工程师、总工程师，主要研究方向为 IPv6、网络安全、大数据。

耿光刚（1980-），男，山东泰安人，博士，中国互联网络信息中心研究员，主要研究方向为机器学习、大数据分析和互联网基础资源安全。