网络钓鱼识别研究综述

付溪，李晖，赵兴文
（西安电子科技大学网络与信息安全学院，陕西 西安 710126）

摘　要：随着互联网的不断发展，网络钓鱼给人们日常生活带来的威胁与日俱增。网络钓鱼识别技术是对抗钓鱼攻击的核心安全技术，可以帮助人们有效避免钓鱼攻击引起的安全威胁。首先，从网络钓鱼的基本概念入手，详细分析了网络钓鱼识别技术的研究现状，然后，对目前网络钓鱼识别的应用场景进行了归纳和总结，最后，对今后可能的研究方向进行了讨论。

关键词：网络钓鱼识别；机器学习；文本分类；特征选择
中图分类号：TN915
文献标识码：A
doi: 10.11959/j.issn.2096−109x.2020062

Survey on phishing detection research

FU Xi, LI Hui, ZHAO Xingwen
School of Cyber Engineering, Xidian University, Xi'an 710126, China

Abstract: With the continuous development of the internet, the threat posed by phishing to people's daily lives is increasing. As a core security technology against phishing attacks, phishing detection technology can help people effectively avoid security threats caused by phishing attacks. Firstly, starting with the basic concepts of phishing, the current application scenarios of phishing identification were summarized. Then, the research status of phishing identification technology in detail were analyzed. Finally, possible future research directions were discussed.

Key words: phishing detection, machine learning, text classification, feature selection

1　引言

近年来，互联网的快速发展极大地丰富了人们的物质文化生活，推动了信息社会的不断进步，但同时吸引了大量非法攻击者进行网络钓鱼攻击。RSA 公司的一份报告显示，2016 年的网络钓鱼事件对全球国际组织造成的间接经济损失高达 90 亿美元[1]。国际反网络钓鱼工作组（APWG）在 2019 年第三季度检测到的网络钓鱼站点总数为 266 387 个，与第二季度相比增长了 46%，达到自

收稿日期：2019−12−05；**修回日期：**2020−02−16
通信作者：赵兴文，xwzhao@xidian.edu.cn
基金项目：国家重点研发计划（2016YFB0801001）；移动互联网陕西省创新团队项目（2018TD-007）；西安市科技计划项目（201809168CX9JC10）
Foundation Item: The National Key R&D Program of China (2016YFB0801001), Mobile Internet Shaanxi Innovation Team Project (2018TD-007), Xi'an Science and Technology Plan Project (201809168CX9JC10)

论文引用格式：付溪, 李晖, 赵兴文. 网络钓鱼识别研究综述[J]. 网络与信息安全学报, 2020, 6(5): 1-10.
FU X, LI H, ZHAO X W. Survey on phishing detection research[J]. Chinese Journal of Network and Information Security, 2020, 6(5): 1-10.

2016 年以来的最高水平[2]。因此，如何及时高效地对网络钓鱼行为进行识别成为亟待解决的安全问题。

网络钓鱼攻击通常会创建一个与合法网站相似度很高的虚假网站，从而通过诱骗用户访问虚假网站来窃取重要隐私信息（如用户姓名、电话、账号密码等）。这将会造成严重的隐私泄露问题，进一步导致用户财产受到威胁。为了应对非法钓鱼攻击，国内外研究人员针对网络钓鱼识别开展了大量研究。早期，由于钓鱼攻击场景单调、攻击站点相对固定的特点，基于白名单、黑名单的识别方法被广泛应用。此外，随着钓鱼攻击方式和攻击者能力的不断提升，基于启发式、机器学习等方法的网络钓鱼识别技术逐渐成为研究重点[3-5]。

本文对网络钓鱼识别技术的应用场景进行了全面的归纳和总结，详细分析了网络钓鱼识别技术的研究现状并对今后可能的研究方向进行了讨论。

2　网络钓鱼识别技术

2.1　基于列表的识别技术

基于列表的识别技术大多通过建立网站的黑白名单实现。黑名单和白名单技术最初被广泛应用于网络钓鱼识别。此技术在将网页呈现给用户之前，首先检查当前网页的 URL 是否与黑名单列表或白名单列表中的某些项相匹配。目前，主流的浏览器集成了黑/白名单以抵御网络钓鱼攻击。例如，谷歌浏览器通过自动更新黑名单列表来阻止恶意站点进行钓鱼攻击，用户可通过 Google 安全浏览 API[6]检查待访问站点的安全性。

白名单方法维护已知安全网站列表来抵御网络钓鱼攻击，只有出现在列表中的网站才被视为可靠的网站。通常的白名单工具需要动态更新通用白名单列表来保证其可用性。但针对个人用户，这会造成一定的冗余，带来较大的维护成本。因此，Han 等[7]提出了一种维护个人白名单的方法。该方法只需通过维护用户个人常用合法网站列表来抵抗网络钓鱼攻击。

黑名单方法主要通过维护已知网络钓鱼站点的列表及对当前访问网站进行检查来实现网络钓鱼攻击的识别。黑名单通常由多个数据源（如垃圾邮件过滤器，用户提交或第三方编制的经过验证的网络钓鱼站点）收集的恶意网站列表构成。Prakash 等[8]使用近似匹配算法将 URL 划分为多个与黑名单条目相匹配的组件，此算法可得到 3%的假阳性率（FPR，false positive rate）和 5%假阴性率（FNR，false negative rate）。Zhang 等[9]设计了一个可以给集中式日志共享设施提供个人定制黑名单的系统。

基于列表的网络钓鱼识别易于实现，具有运行速度高和 FPR 低的优点。但由于待检测 URL 与列表的极小差异都会导致匹配失败，攻击者往往通过修改钓鱼网页的 URL 来规避检测，这会导致 FNR 增加。与此同时，统计数据显示[10]，63%的网络钓鱼网站的生命周期仅为 2 h，但 47%～83%的网络钓鱼网站在 12 h 后才会被添加到黑名单中。由于网络钓鱼网站产生速度快、生命周期短，因此黑名单需要从其来源频繁更新，但这样会导致系统资源占用率较大，且基于列表的系统无法检测零日钓鱼攻击。因此，仅使用基于列表的网络钓鱼网站的识别技术无法满足现有的网络安全需求，研究者往往使用基于列表的识别方法与基于机器学习的识别方法相结合加快检测速度。

2.2　基于启发式的识别技术

基于启发式的技术不依赖任何预编译列表，此方法通常需要人工提取网页中的启发式规则来识别网络钓鱼。目前存在许多基于启发式的方法[11-18]，这些方法主要根据可疑网页的 URL、文档对象模型（HTML DOM, hyper text markup language document object model）和第三方服务来识别网络钓鱼站点。

Gastellier-Prevost 等[11]开发了一种基于 HTML DOM 的方法 Phishark。此方法定义了识别网络钓鱼网站的 20 种启发式规则，并对各项规则的有效性进行了评估。很多研究[12-17]专注于利用搜索引擎结果来识别网络钓鱼网站，Zhang 等[18]提出了一种基于内容识别网络钓鱼的方法 CANTINA，此方法使用词频－逆文档频率（TF-IDF，term frequency-inverse document frequency）算法提取页面标题和内容中出现频率最高的关键词，并通过 Google 等主流搜索引擎对关

键词进行检索，将出现在前 N 个搜索结果中的待识别网站视为合法网站。基于搜索引擎技术的缺点是无法识别出托管在受感染服务器上的网络钓鱼站点。当服务器受到感染时，使用搜索引擎的识别方法很有可能将其托管的钓鱼网站识别为无害网站，从而导致识别的 FNR 较高。另外，由于新注册的合法站点缺少搜索结果，因而，很有可能将无害网站识别为钓鱼网站，导致识别的 FPR 很高。因此，Rao 等[19]提出了一个应用程序 Jail-Phish。它可以提高基于搜索引擎技术的准确性，可以识别出托管在受感染服务器上的网络钓鱼站点以及新注册的合法站点，并且准确率可达 98.61%，而 FPR 小于 0.64%。

基于启发式的方法不需要任何预编译列表，只需通过提取启发式特征来识别网络钓鱼网站，这些方法可以检测零日钓鱼攻击。但是，启发式规则往往比较简单，容易被攻击者规避，而且此方法具有语言依赖性，从网页的文本内容中提取的某些文本特征不能用于识别不同语言类型的网络钓鱼网站。同时，这些方法在识别仅由图像和脚本等嵌入对象组成的仿冒网页方面效果较差。目前，基于启发式的识别方法往往作为钓鱼识别系统的预处理部分提升系统速度。

2.3　基于机器学习的识别技术

基于机器学习（ML）的识别技术依据从网站中提取的特征识别网络钓鱼。通常网络钓鱼网站与合法网站存在可区分的特征，并且机器学习在这方面可以进行有效处理，因此机器学习方法被广泛应用于网络钓鱼网站的识别。基于机器学习识别网络钓鱼的解决方案主要依据来自网站的 URL、源代码和证书等特征将网页识别为钓鱼网站或合法网站。

2.3.1　基于 URL 的特征提取

URL 特征主要包括词汇特征和外部特征[20-21]，如图 1 所示。词汇特征一般可以从 URL 字符串中快速提取，外部特征则需要查询远程服务器。有研究[20]表明，仅使用词汇特征的分类器与使用全部特征（词汇+外部）的分类器性能相当。因此，词汇特征更适合在客户端识别网络钓鱼。

Yuan 等[22]提出了从 URL 和网页链接中提取特征以识别出网络钓鱼网站及其目标的方法，并比较了许多机器学习算法用于识别网络钓鱼的性能，其中深森林（DF，deep forest）算法效果最好，可实现 98.3%的 FNR 和 2.6%的 FPR。Garera 等[23]提出了发起钓鱼攻击的 4 种不同类别的 URL 混淆技术，并提出了 18 种识别网络钓鱼 URL 的特征。Mc-Grath 等[24]对网络钓鱼 URL 和非网络钓鱼 URL 进行了比较分析，发现网络钓鱼与非网络钓鱼的 URL 长度和域长度区别较大。因此，基于 URL 长度和域长度的特征已成功用于后续网络钓鱼识别研究中。Li 等[25]提出了将 URL 和 HTML 特征用于混合分类器的方法，该方法利用不同的模型的互补性将 GBDT、XGBoost 和 LightGBM 组合建立分类器，最终可到达 97.3%的准确率。最近，研究人员提出了将深度神经网络用于恶意 URL 识别系统进行特征提取和分类的方法[26-28]。

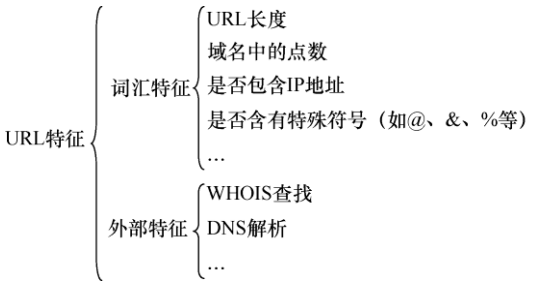


图 1　URL 特征
Figure 1　URL feature

2.3.2　基于源代码的特征提取

基于源代码的技术从可疑 URL 的源代码中提取特征（如基于超链接、基于文本和基于图像的特征等）。基于超链接的特征[29-31]包括外部链接、断开链接和空链接等。基于文本的特征[18,32-34]主要从网页中提取关键字（如 TF-IDF 方法）。基于图像的特征包括提取待检测网站的徽标[12,35-38]、网页的屏幕截图[39]等，以识别钓鱼网站。近年来，Fang 等[40]针对攻击者将恶意 JavaScript 注入网页以进行网络钓鱼和获取机密信息的问题，提出了一种基于长短期记忆（LSTM，long short-term memory）的恶意 JavaScript 检测模型，该模型从字节码的语义层提取特征。实验表明，基于 LSTM 的检测模型可以区分恶意 JavaScript 代码并达到 99.51%的准确率。

2.3.3　基于证书的特征提取

近年来，部分学者对于能否将网站证书信息

应用于网络钓鱼对抗进行了探索。

Dong 等[41]提出了从证书中提取特征（如证书有效期和发行人字段）进行钓鱼网站识别的方法，最佳可实现 95%的识别精度。Mensah 等[42]尝试仅使用从证书和握手信息中提取的特征区分钓鱼网站和合法网站，但得出的结论是，不可能仅使用此类信息来区分网站类别。Drury 等[43]对网站证书中包含的信息能否区分网络钓鱼与良性网站进行了探索，得出了良性网站和钓鱼网站的证书之间没有明显差异的结论。但 Drury 等进行了进一步的研究，直接将网络钓鱼网站的证书与其目标证书进行比较，发现虽然钓鱼网站和合法网站的证书没有明显区别。但是，对于 15 个最受欢迎的网站，其网络钓鱼证书明显与其目标网站证书不同，特别是证书中颁发者和主题信息之间存在明显差异。

目前，大多数基于机器学习的方法将以上 3 种类型组合使用，如 Rao 等[30]提出了从 URL、源代码和第三方服务中提取特征的分类模型，最高可达到 99.55%准确度。胡向东等[44]提出了一种结合页面敏感文本特征和 logo 图像特征的金融类钓鱼网页检测方法，并取得 97%的召回率。方勇等[45]提出了一种使用 LSTM 和随机森林的混合算法模型，该模型使用 LSTM 网络提取字符序列特征，并结合传统的非字符序列特征构建实验特征集，弥补了 LSTM 算法特征单一的问题，模型准确率达到 98.52%。

基于机器学习的识别方法可识别零日钓鱼攻击，同时具有高准确率与可扩充性的优点，模型修正较为简单（在原有数据集中加入新的钓鱼数据即可）。但其识别性能与数据样本的分布密切相关。当钓鱼样本与合法网站分布差异较大时，识别性能会受到很大影响。同时，由于用于识别网络钓鱼的特征种类繁多，规模庞大，往往导致冗余特征的存在，使技术的成本（存储、训练时间等）增加。因此，为了降低系统计算开销、保证系统效率、提高识别准确率，部分学者[46-47]对于将特征选择技术用于基于机器学习的钓鱼识别方法进行了尝试。

3　网络钓鱼检测的应用场景

3.1　邮件钓鱼识别

电子邮件是网络钓鱼中最常用且最具影响力的手段。据 APWG 统计，网络钓鱼电子邮件的数量从 2014 年到 2015 年增长了近 1 倍，2017 年 1 月至 2017 年 6 月网络钓鱼电子邮件的数量约为 100 000[48-49]。微软的一份报告显示从 2018 年 1 月到 2018 年 12 月，网络钓鱼电子邮件数量增长了 250%[50]。由于具有增长迅速且危害巨大的特点，电子邮件钓鱼受到了研究人员的广泛关注[51]。

在之前的研究中，只有部分研究可以处理零日钓鱼攻击，且迄今为止公开的方法具有较高的 FPR 和较低的准确性。近年来，研究者对上述问题进行了进一步探索。Gascon 等[52]提出了一种基于发件人个人资料识别钓鱼邮件的检测方法，该方法可识别 90%以上的钓鱼邮件。Smadi 等[53]提出了一种将神经网络与强化学习相结合的新型框架在线检测网络钓鱼攻击，该框架可以处理零日网络钓鱼攻击，并具有高准确率（98.63%）和较低 FPR（1.81%）的特点。Fang 等[54]提出了一种新的网络钓鱼电子邮件检测模型 THEMIS，该模型用于在电子邮件标题、正文和字符级别对电子邮件建模，该模型可达到 99.848%的准确率及 0.043%的 FPR。高一男等[55]提出了一种基于大数据平台 Hadoop 识别钓鱼邮件的方法。此方法采用 HDFS 存储数据集，MapReduce 进行并行计算，Mahout 的贝叶斯算法识别钓鱼邮件，取得了良好的效果。

3.2　社交网站钓鱼识别

近年来，钓鱼攻击者将目标转向在线社交网络（OSN）用户，如 Twitter、Facebook、Myspace 等。由于 Twitter 信息传播的广泛性以及相比电子邮件更难以被反钓鱼组织检测的特点，网络钓鱼者开始使用其作为传播网络钓鱼的媒介。因此，检测推文是否含有网络钓鱼 URL 并实时向用户发送提醒是至关重要的。

Aggarwal 等[56]提出使用 Twitter 的特有特征（如推文内容、推文长度、推文数量和点赞者比率等）结合 URL 特征来识别网络钓鱼的模型，该模型可达到 92.52%的准确性。Sharma 等[57]在基于机器学习的网络钓鱼检测实验中使用了基于 URL、基于推文、基于 WHOIS、基于用户和基于网络的特征，并最终得出在分类中添加更多特征集（尤其是基于推文的特征）会显著提高网络钓

鱼检测性能的结论。Liew 等[58]提出了一种基于随机森林（RF，random forest）算法的预警机制，最终在 11 个最佳分类特征上达到 94.75%的准确率，高于其他相关研究中 94.56%的准确率。

3.3　移动设备中的网络钓鱼识别

随着移动技术和互联网的飞速发展，针对移动网络的攻击和欺诈行为不断发生。在线购物、金融和社交应用用户正面临移动网络钓鱼的新威胁[59]。目前，有许多针对 PC 端网络钓鱼攻击的防御和保护方案，但并没有针对移动平台的有效安全措施[60-61]，大多数用于 PC 设备上的反网络钓鱼工具无法有效地解决移动设备上的网络钓鱼攻击问题[62]。移动平台是网络钓鱼攻击的新目标。

Virvilis 等[63]表明 Andriod 默认浏览器不提供任何形式的网络钓鱼防护功能。段青[64] 在不同机型系统中进行了钓鱼攻击测试，结果表明现有安全软件无法对钓鱼攻击做出识别。为了有效检测 Android 平台上的网络钓鱼攻击，Ndibwile 等[65]提出了一种移动应用程序原型 UnPhishMe。UnPhishMe 通过检测当前登录界面在进行身份验证后是否跳转到其他网页，确定网站是否为钓鱼网站。Liu 等[66]提出了一种基于改进朴素贝叶斯算法的框架，此框架通过评估不同属性的权重，调整钓鱼网站与合法网站的概率比的方法，提高检测的准确率。刘永明等[67]提出了一种利用图像相似性识别 Andriod 钓鱼应用的方法，该方法通过动态截取 Andriod 应用的人机交互界面，计算其与目标应用界面的图像相似度来识别钓鱼应用。实验表明，该方法可以有效检测 Android 平台上的恶意钓鱼应用程序。

4　识别方法小结

本文第 2 节和第 3 节主要介绍了几类网络钓鱼识别技术和现有的研究场景，并结合已有的研究成果对其进行分析。表 1 从识别方法及评价指标等方面对这几类识别方法进行了比较和总结。表 2 从算法、数据集、准确率等方面对几种典型的基于机器学习的方案进行了对比。

由表 1 可知，基于列表的识别方法[1,8]检测速度快，实时性高，但黑白名单需要从其来源频繁更新，且无法检测零日钓鱼攻击。基于启发式的

表 1　典型识别方法比较
Table 1　Comparison of typical detection methods

典型工作	识别方法			评价指标			
	基于列表	启发式	机器学习	准确率	FNR	FPR	时间
Google API[6]	√			—	—	—	0.08 s
PhishNet[8]	√	√		—	3%	5%	0.001 s
CANTINA[18]		√		95%	—	6%	—
Phishark[11]		√		—	—	2.4%	2 s
Website logo method[12]		√		93.4%	0.2%	13%	—
SHLR[68]		√	√	98.9%	—	—	0.029 ms
URLs based method[22]			√	98.3%	—	2.6%	—
CGRU[27]			√	99.6%	—	—	640.78 s
MFPD[28]			√	98.99%	—	0.59%	—
PEDS[53]			√	98.63%	0.93%	1.81%	—
THEMIS[54]			√	99.85%	0.128%	0.043%	—
PhishAri[56]			√	92.52%	7.78%	9.6%	0.425 s
Twitter phishing[58]			√	94.75%	5.12%	5.3%	—
HEFS[46]			√	94.6%	—	—	0.052 ms

注：空白表示不适用，“—”表示无法获得。

表 2　基于机器学习的方案比较
Table 2　Comparison of machine learning-based solutions

典型方法	方法	算法	数据集大小	特征	准确率
URLs based method[22]	ML	DF	6 197	12 个 URL 特征	97.7%
		KNN			90.3%
		RF			96.0%
		LR			95.4%
CGRU[27]	DL	卷积神经网络+门控递归单元	407 212	从原始 URL 得到语义特征	99.6%
A stacking model[25]	DL	GBDT+XGBoost+LightGBM	49 947	8 个 URL 特征+12 个 HTML 特征	97.30%
			51 103		98.60%
MFPD[28]	DL	LSTM+ XGBoost	2 010 779	24 个 URL 特征+文本特征	99.41%
PEDS[53]	ML	强化学习	12 326	从电子邮件提取 50 个特征	98.6%
THEMIS[54]	DL	RCNN	7 781	从电子邮件的文本内容得到语义特征	99.84%
IPDPS[69]	ML	ANFIS	11 056	22 个文本特征+5 个图像特征+8 个框架特征	98.3%
HEFS[46]	ML	RF	10 000	从 URL 和 HTML 源代码提取 48 个特征	96.17%
		C4.5			94.37%
		SVM			92.20%
		NB			84.10%

注：空白表示不适用。

方法[12-18]不需要任何预编译列表，相较于基于列表的检测方法，基于启发式的检测方法 FPR 较高，由于规则更新依赖于专业领域知识，规则更新存在一定困难。基于机器学习的方法[27-28, 53-54]可以获得更高的准确率和较低的 FPR，但由于机器学习算法需要对大量数据建立分类器，分类器的建立需要大量时间，系统实时性低。为了增强系统的实时性，Chiew 等[46]对于基于机器学习的网络钓鱼特征选择进行了探索，设计了一个具备特征子集选取功能的 HEFS 系统，但没有给出该系统具体的 FNR、FPR 测试结果，因此无法确认其方法的实用性。Ding 等[68]提出了一种将启发式规则与机器学习方法结合的混合方法，该方法可以提高系统的实时性，因此，钓鱼方法融合可作为未来研究方向之一。另外，一些基于机器学习的方法对于 Twitter 钓鱼攻击[56,58]检测的准确率较低，FPR 较高，仍有研究的空间。

由表 2 可知，在准确率方面，机器学习方法中以 RF 算法为代表的集成学习方法相比单一算法（如 KNN、C4.5、SVM、NB 等）可获得更高的准确率。深度学习（DL）方法可以获得比机器学习算法更高的准确率。在特征方面，基于机器学习方法的特征往往需要研究人员人工提取，但基于深度学习的方法避免了手动提取特征，可以实现自动提取。基于深度学习的方法往往需要比机器学习更大的训练数据集，而且分类器训练的时间更长，系统实时性不佳。

5　可研究方向

5.1　新颖特征发掘

现有的反网络钓鱼方法仅针对特定类型的攻击。一方面，互联网的发展，特别是移动网络钓鱼、语音网络钓鱼和社交媒体网络钓鱼的出现，对网络钓鱼识别的准确率和实时性提出了更高要求。另一方面，现有的基于机器学习的零日钓鱼攻击检测方法严重依赖有限类型的特征，网络钓鱼者可以通过学习其特征来逃避检测。为了应对以上问题，研究人员可以从移动网络、语音网络和社交网络的特有特征出发，发掘发现新的、健壮的预测特征，这可能是未来钓鱼网站识别的发展方向之一。

5.2　特征选择

近年来，部分学者[46,69-70]对于基于机器学习的网络钓鱼特征选择进行了进一步探索。基于机器学习的网络钓鱼检测系统的准确性主要取决于所选特征。大多数反网络钓鱼研究者着眼于提出新颖特征或优化分类算法[71-72]，但目前用于识别网络钓鱼的特征种类繁多，可能仍然存在某些不相关的特征，使技术的成本（存储、训练时间等）增加。因此，为了在不降低准确性的前提下降低系统开销、提高检测效率，需要一个特征选择框架来识别真正有效的最小特征集合。要获得最小特征集合，需要选择一个截止等级，仅选择位于该截止等级之前的特征（目前截止等级往往是随意给出的）。因此，截止等级的选取可以作为未来网络钓鱼识别可探索的方向之一。

5.3　数据集建立

由于基于机器学习的网络钓鱼检测方法得到研究者的广泛应用，部分研究者探讨了如何收集和整理机器学习所需的网络钓鱼数据集[73]。一方面，机器学习算法的主要资源是数据集，为了提高机器学习算法的性能，减少研究者的前期数据收集的工作负担，合理的数据集构造是很有必要的；另一方面，构造合理的标注数据集有助于更好地比较与评测不同特征。钓鱼网站的变化十分迅速，导致公开的数据集的实时性受到挑战，与此同时，基于视觉相似性的钓鱼检测方法可有效识别由图片构成的钓鱼网站[74]，但这种识别方法存在网页图片数量庞大、获取困难等问题。因此，未来可通过收集最新的钓鱼网站图像样本构建新鲜数据集方面进行探索。

5.4　钓鱼识别方法融合

近年来，部分学者[68]对钓鱼识别方法的融合进行了探索，不同的识别方法适用于不同场景。例如，基于列表和基于启发式规则的方法检测时间短，但准确率较低；基于机器学习的方法检测准确率高，但耗时较长。不同识别方法之间存在很强的互补性，不同识别方法的融合对于改善整体系统的准确率和实时性有着重要影响。因此，使用基于列表或基于启发式的识别方法作为系统预处理部分快速识别网络钓鱼，然后使用基于机器学习的识别方法提高钓鱼识别准确率，这可能是未来钓鱼网站识别的发展方向之一。

5.5　钓鱼对抗

近年来，部分学者[75]选择特征生成对抗性样本，模拟攻击基于机器学习的分类器。研究表明，基于机器学习分类器模型的样本可以通过更改最多 4 个特征值来绕过分类器，即网络钓鱼检测机制容易受到对抗性学习技术的攻击。机器学习是区分网络钓鱼网站和合法网站的一种有前途的技术，但机器学习方法容易受到对抗性学习的影响，使分类器的准确性降低。因此，选择特征生成对抗性样本从而测试已有钓鱼识别系统的鲁棒性可以作为未来探索方向之一。

5.6　钓鱼目标发现

上述的反网络钓鱼方法都试图识别网络钓鱼网页，但缺乏识别网络钓鱼模仿的合法网页（网络钓鱼目标）的技术。如果不能识别网络钓鱼目标，任何反网络钓鱼技术都会变得不完整，近年来部分学者对于钓鱼目标发现进行了探索[22,32]。由于网络钓鱼目标在确认合法网页上是否存在网络钓鱼攻击方面起着至关重要的作用，查找网络钓鱼目标有利于分析攻击行为，并帮助用户导航到合法网页。有时网络钓鱼者使用的伪装技术，使钓鱼目标难以识别。例如，如果网络钓鱼者仅使用图像、脚本等嵌入对象创建网页，则使用现有方法识别目标变得困难。因此，发现钓鱼目标，提醒钓鱼目标所有者采取必要的对策并帮助用户导航到目标合法网页，可作未来的研究方向之一。

6　结束语

网络钓鱼识别作为网络安全领域的热点问题引起了研究人员的广泛关注。针对这一问题，本文首先介绍了网络钓鱼的研究背景和基本概念，并阐述了网络钓鱼识别的研究技术以及应用场景，此外，针对现有的钓鱼识别方案进行了详细的分析，从识别方法及评价指标等方面进行了对比，最后对今后可能的研究方向进行了讨论。

参考文献：

[1]　BLEAU H. 2017 Global fraud and cybercrime forecast[EB].

[2] Phishing activity trends report 3rd quarter 2019[R].

[3] 王惟. 反钓鱼技术综述[J]. 山东广播电视大学学报, 2011(3): 45-46, 49.
WANG W. Review of anti-phishing technology[J]. Journal of Shandong Radio and TV University, 2011(3): 45-46, 49.

[4] 李江丰, 王玮. 钓鱼网站的识别与分析方法研究[J]. 通信管理与技术, 2018(3): 62-64.
LI J F, WANG W. Research on recognition and analysis methods of phishing websites[J]. Communications Management and Technology, 2018(3): 62-64

[5] 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.

[6] Google safe browsing APIs[EB].

[7] HAN W, CAO Y, BERTINO E, et al. Using automated individual white-list to protect web digital identities[J]. Expert Systems with Applications, 2012, 39(15).

[8] PRAKASH P, KUMAR M, KOMPELLA R R, et al. PhishNet: predictive blacklisting to detect phishing attacks[C]//29th INFOCOM 2010.

[9] ZHANG J, PORRAS P A, ULLRICH J. Highly predictive blacklisting[C]//17th USENIX Security Symposium 2008.

[10] SHENG S, WARDMAN B, WARNER G, et al. An empirical analysis of phishing blacklists[C]//The 6th Conf. Email Anti-Spam (CEAS).

[11] GASTELLIER-PREVOST S, GRANADILLO G G, LAURENT M. Decisive heuristics to differentiate legitimate from phishing sites[C]//Network & Information Systems Security. 2011.

[12] KANG L C, CHANG E H, SZE S N, et al. Utilisation of website logo for phishing detection[J]. Computers & Security, 2015(54): 16-26.

[13] DUNLOP M, GROAT S, SHELLY D. GoldPhish: using images for content-based phishing analysis[C]//Fifth International Conference on Internet Monitoring & Protection. 2010.

[14] HUH J H, KIM H. Phishing detection with popular search engines: simple and effective[M]//Foundations and Practice of Security. Berlin Heidelberg: Springer, 2011.

[15] JAIN A K, GUPTA B B. Two-level authentication approach to protect from phishing attacks in real time[J]. Ambient Intelligence and Humanized Computing, 2018, 9(6): 1783-1796.

[16] TAN C L, KANG L C, WONG K S, et al. PhishWHO: phishing webpage detection via identity keywords extraction and target domain name finder[J]. Decision Support Systems, 2016(88): 18-27.

[17] VARSHNEY G, MISRA M, ATREY P K. Improving the accuracy of search engine based anti-phishing solutions using lightweight features[C]//11th ICITST 2016.

[18] ZHANG Y, HONG J I, CRANOR L F. Cantina: a content-based approach to detecting phishing web sites[C]//16th WWW, 2007.

[19] RAO R S, PAIS A R. Jail-phish: an improved search engine based phishing detection system[J]. Computers & Security, 2019(83): 246-267.

[20] LE A, MARKOPOULOU A, FALOUTSOS M. PhishDef.URL names say it all[J]. CoRR abs/1009.2275, 2010.

[21] MA J, SAUL L K, SAVAGE S, et al. Beyond blacklists: learning to detect malicious Web sites from suspicious URLs[C]//KDD. 2009.

[22] YUAN H, CHEN X,LI Y, et al. Detecting phishing websites and targets based on URLs and webpage links[C]//24th ICPR 2018.

[23] GARERA S, PROVOS N, CHEW M, et al. A framework for detection and measurement of phishing attacks[C]//ACM Workshop on Recurring Malcode. 2007.

[24] MC-GRATH D K, GUPTA M. Behind phishing: an examination of phisher modi operandi[C]//5th NSDI 2008.

[25] LI Y, YANG Z, XU C, et al. A stacking model using URL and HTML features for phishing webpage detection[J]. Future Generation Comp Syst, 2019(94): 27-39.

[26] SAHINGOZ O K, BUBER E, DEMIR O, et al. Machine learning based phishing detection from URLs[J]. Expert Syst Appl, 2019(117): 345-357.

[27] YANG W, ZUO W, CUI B. Detecting malicious URLs via a keyword-based convolutional gated-recurrent-unit neural network[J]. IEEE Access, 2019(7): 29891-29900.

[28] YANG P, ZHAO G, ZENG P. Phishing website detection based on multidimensional features driven by deep learning[J]. IEEE Access, 2019(7): 15196-15209.

[29] MARCHAL S, ARMANO G, GRONDAHL T, et al. Off-the-hook: an efficient and usable client-side phishing prevention application[J]. IEEE Trans. Computers, 2017, 66(10): 1717-1733.

[30] RAO R S, PAIS A R. Detection of phishing websites using an efficient feature-based machine learning framework[J]. Neural Computing and Applications, 2019, 31(8): 3851-3873.

[31] SHIRAZI H, BEZAWADA B, RAY I. Unbiased phishing detection using domain name based features.[C]//SACMAT. 2018.

[32] MARCHAL S, SAARI K, SINGH N , et al. Know your phish: novel techniques for detecting phishing sites and their targets[C]//36th ICDCS. 2016.

[33] 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.

[34] XIANG G, HONG J I, ROSÉ C P, et al. CANTINA+: a feature-rich machine learning framework for detecting phishing web sites[J]. ACM Trans Inf Syst Secur, 2011, 14(2): 21:1-21:28.

[35] FATT J C S, LENG C K, NAH S S. Phishdentity: leverage website favicon to offset polymorphic phishing website[C]//ARES. 2014.

[36] CHIEW K L, CHOO J S F, SZE S N, et al. Leverage website favicon to detect phishing websites[J]. Security and Communication

Networks, 2018: 11.

[37] 周诚诚, 张代远. 利用图像识别技术过滤海量可疑钓鱼网站[J]. 计算机技术与发展, 2012(11): 246-249.
ZHOU C C, ZHANG D Y. Using image recognition technology to filter mass suspicious phishing sites[J]. Computer Technology and Development, 2012(11): 246-249.

[38] 肖洪云. 图像识别在钓鱼检测中的应用[J]. 沧州师范学院学报, 2012(3): 75-79.
XIAO H Y. The application of image recognition to the fishing detection[J]. Journal of Cangzhou Normal University, 2012(3): 75-79.

[39] HARA, YAMADA, MIYAKE. Visual similarity-based phishing detection without victim site information[C]//ICICS. 2009.

[40] FANG Y, HUANG C, LIU L, et al. Research on malicious JavaScript detection technology based on LSTM[J]. IEEE Access, 2018, 6: 59118-59125.

[41] DONG Z, KAPADIA A, BLYTHE J, et al. Beyond the lock icon: real-time detection of phishing websites using public key certificates[C]//eCrime. 2015.

[42] MENSAH P, BLANC G, OKADA K, et al. AJNA: anti-phishing JS-based visual analysis, to mitigate users' excessive trust in SSL/TLS[C]// 4th BADGERS@RAID. 2015.

[43] DRURY V, MEYER U. Certified phishing: taking a look at public key certificates of phishing websites[C]//15th SOUPS @ USENIX Security Symposium. 2019.

[44] 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, 2017(2): 35-42.
HU X D, LIU K, ZHANG F, et al. Financial phishing detection method based on sensitive characteristics of webpage[J]. Chinese Journal of Network and Information Security, 2017(2): 35-42.

[45] 方勇, 龙啸, 黄诚, 等. 基于 LSTM 与随机森林混合构架的钓鱼网站识别研究[J]. 四川大学学报（工程科学版）, 2018(5).
FANG Y, LONG X, HUANG C, et al. Research on classifying phishing URLS using hybrid architecture of LSTM and random forest[J]. Advanced Engineering Sciences, 2018(5).

[46] CHIEW K L, TAN C L, WONG K, et al. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system[J]. Inf Sci, 2019, 484: 153-166.

[47] BABAGOLI M, AGHABABA M P, SOLOUK V. Heuristic nonlinear regression strategy for detecting phishing websites[J]. Soft Comput, 2019, 23(12): 4315-4327 .

[48] Phishing activity trends report 4th quarter 2016[EB].

[49] Phishing activity trends report 1st-3rd quarter 2015[EB].

[50] Microsoft security intelligence report[R]. 2018.

[51] 王晓丽. 钓鱼邮件攻击防范指南[J]. 计算机与网络, 2018, 581(13): 56-57.
WANG X L. Guidelines for preventing phishing email attacks[J]. Computer & Network2018, 581(13): 56-57.

[52] GASCON H, ULLRICH S, STRITTER B, et al. Reading between the lines: content-agnostic detection of spear-phishing emails[C]//21st RAID. 2018.

[53] SMADI S, ASLAM N, ZHANG L. Detection of online phishing email using dynamic evolving neural network based on reinforcement learning[J]. Decision Support Systems, 2018, 107: 88-102.

[54] FANG Y, ZHANG C, HUANG C, et al. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism.[J]. IEEE Access, 2019,7: 56329-56340.

[55] 高一男, 蔡满春. 基于 Hadoop 和 Mahout 的钓鱼邮件检测技术研究[J]. 电脑知识与技术, 2016, 12(11): 27-30.
GAO Y N,CAI M C. Research of phishing-mail detection based on Hadoop and Mahout[J]. Computer Knowledge and Technology, 2016, 12(11): 27-30.

[56] AGGARWAL A, RAJADESINGAN A, KUMARAGURU P. PhishAri: automatic realtime phishing detection on Twitter[J]. CoRR abs/1301.6899, 2013.

[57] SHARMA N, SHARMA N, TIWARI V, et al. Real-time detection of phishing Tweets[C]//The Fourth International Conference on Computer Science, Engineering and Applications. 2014.

[58] LIEW S W, SANI N F M, ABDULLAH M T, et al. An effective security alert mechanism for real-time phishing tweet detection on twitter[J]. Computers & Security, 2019.

[59] 石春宏. 移动终端如何识别钓鱼手段与防范[J]. 电脑知识与技术, 2017(31): 43-44.
SHI C H. How to identify and prevent phishing on mobile terminals[J]. Computer Knowledge and Technology, 2017(31): 43-44.

[60] BICAKCI K, UNAL D, ASCIOGLU N, et al. Mobile authentication secure against man-in-the-middle attacks[C]//2nd Mobile Cloud 2014.

[61] GOEL D, JAIN A K. Mobile phishing attacks and defence mechanisms: state of art and open research challenges[J]. Computers & Security, 2018, 73: 519-544.

[62] WU L, DU X, WU J. Effective defense schemes for phishing attacks on mobile computing platforms[J]. IEEE Trans Vehicular Technology. 2016, 65(8): 6678-6691.

[63] VIRVILIS N, TSALIS N, MYLONAS A, et al. Mobile devices: a phisher′s paradise[C]//11th International Conference on Security and Cryptography. 2014.

[64] 段青. 一种移动平台钓鱼攻击的解决方法[J]. 信息安全与技术. 2016, 7(4):50-54.
DUAN Q. A solution for phishing attackon Android platform[J]. Cyberspace Security, 2016, 7(4):50-54.

[65] NDIBWILE J D, KADOBAYASHI Y, FALL D. UnPhishMe: phishing attack detection by deceptive login simulation through an android mobile App[C]//12th AsiaJCIS. 2017.

[66] LIU D, LIU D, LI Y, et al. Efficient Android phishing detection based on improved naïve bayes algorithm[C]//10th ICSI. 2019.

[67] 刘永明, 杨婧. 基于图像相似性的 Android 钓鱼恶意应用检测方

法[J]. 计算机系统应用, 2014, 23(12):170-175.
LIU Y M, YANG J. Detection of android phishing malwares based on image similarity[J]. Computer Systems & Applications, 2014, 23(12):170-175.

[68] DING Y, LUKTARHAN N, LI K, et al. A keyword-based combination approach for detecting phishing webpages[J]. Computers & Security, 2019, 84: 256-275.

[69] ADEBOWALE M A, LWIN K T, SÁ NCHEZ E, et al. Intelligent web-phishing detection and protection scheme using integrated features of images, frames and text[J]. Expert Syst Appl, 2019, 115: 300-313.

[70] ABUTAIR H, BELGHITH A, ALAHMADI S. CBR-PDS: a case-based reasoning phishing detection system[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(7): 2593-2606.

[71] HE M, HORNG S J, FAN P, et al. An efficient phishing webpage detector[J]. Expert Syst Appl, 2011, 38(10): 12018-12027.

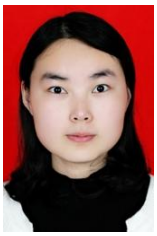
[72] MOHAMMAD R M, THABTAH F, MCCLUSKEY L. An assessment of features related to phishing websites using an automated technique[C]//7th ICITST. 2012.

[73] FALCONIERI V. Open dataset of phishing and tor hidden services screen-captures[J]. CoRR abs/1908.02449, 2019.

[74] 张茜, 延志伟, 李洪涛, 等. 网络钓鱼欺诈检测技术研究[J]. 网络与信息安全学报, 2017, 3(7): 11-28.
ZHANG X, YAN Z W, LI H T. et al. Research of phishing detection technology[J]. Chinese Journal of Network and Information Security, 2017, 3(7): 11-28.

[75] SHIRAZI H, BEZAWADA B, RAY I, et al. Adversarial sampling attacks against phishing detection[M]//Data and Applications Security and Privacy. Berlin: Springer. 2019.

[作者简介]



付溪（1996– ），女，陕西西安人，西安电子科技大学硕士生，主要研究方向为信息安全。

李晖（1968– ），男，河南灵宝人，博士，西安电子科技大学教授、博士生导师，主要研究方向为密码学、无线网络安全、云计算安全、信息论与编码理论。

赵兴文（1977– ），男，广西玉林人，博士，西安电子科技大学副教授，主要研究方向为人工智能在网络安全中的应用、多方参与的数据安全共享、匿名认证等保护隐私的密码协议。