网络钓鱼检测研究综述

谢丽霞[1]，张 浩[2]，杨宏宇[2*]，胡 泽[2]，成 翔[3,4]，张 良[5]

（1. 中国民航大学 计算机科学与技术学院，天津 300300；2. 中国民航大学 安全科学与工程学院，天津 300300；3. 扬州大学 信息工程学院，扬州 225127；4. 中国民航大学 民航信息安全评估中心，天津 300300；5. 亚利桑那大学 信息学院，图森 AZ85721）

摘要 网络钓鱼作为一种社会工程攻击手段，旨在通过伪装成可信任的实体，如银行、社交媒体平台或政府机构，通过虚假的电子邮件、网站或消息来欺骗受害者。研究者主要通过各种技术手段检测网络钓鱼攻击，但当前检测研究仍面临三方面问题。1）攻击者采用伪装、漏洞利用和规避技术以逃避检测。2）现有的检测方法存在可解释性差、实时性低以及概念漂移等问题。3）由于缺乏足够的可解释性，造成用户对检测结果不信任。该文从应用场景、数据集、检测方法等方面对当前检测研究进行归纳与总结，并提出当前面临的问题以及展望未来可能的研究热点。

关键词 网络钓鱼；网络钓鱼检测；深度学习；机器学习

中图分类号 TP393 **文献标志码** A DOI 10.12178/1001-0548.2023273

A Review of Phishing Detection Research

XIE Lixia[1], ZHANG Hao[2], YANG Hongyu[2*], HU Ze[2], CHENG Xiang[3,4], and ZHANG Liang[5]

(1. School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China; 2. School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China; 3. School of Information Engineering, Yangzhou University, Yangzhou 225127, China; 4. Information Security Evaluation Center of Civil Aviation, Civil Aviation University of China, Tianjin 300300, China; 5. School of Information, University of Arizona, Tucson AZ85721, USA)

Abstract Phishing, as a form of social engineering attack, aims to deceive victims by masquerading as a trustworthy entity such as a bank, social media platform, or government agency, using false emails, websites, or messages. Researchers primarily employ various technological means to detect phishing attacks, yet current detection studies face three main challenges. Firstly, attackers employ disguise, exploit vulnerabilities, and employ evasion techniques to evade detection. Secondly, existing detection methods suffer from poor interpretability, low real-time capabilities, and issues like concept drift. Lastly, due to insufficient interpretability, users may lack trust in the detection results. This paper summarizes the current detection researches from the aspects of application scenarios, datasets, detection methods, etc., and puts forward the current problems and prospects the possible research hotspots in the future.

Key words phishing; phishing detection; deep learning; machine learning

随着信息技术的迅猛发展，人类社会已进入高度数字化的时代。然而，这也伴随着网络安全威胁不断增加的问题。其中，网络钓鱼作为一种极具欺骗性和破坏性的威胁正日益严重。网络钓鱼运用社会工程学的手段，旨在伪装成合法实体或机构，引诱个人透露敏感信息、点击恶意链接或执行欺诈行为。其多样化的形式和不断升级的技术，使其已经成为网络犯罪中的一个关键组成部分。它不仅对个人隐私构成直接威胁，也对商业、政府和社会稳定产生严重影响。根据 2022 年第 4 季度的 APWG（Anti-Phishing Working Group）发布的网络钓鱼趋势报告[1]，2022 年全球网络钓鱼攻击达到创纪录的 470 余万次。近 4 年来，网络钓鱼攻击每年增长超过 150%。此外，根据 2023 年 Verizon 发布的数据泄露调查报告[2]，几乎一半的数据泄露事件直接或间接与网络钓鱼有关。

研究者已经采用多种方法来进行网络钓鱼检测，但现有这些方法都存在一定的局限性。如基于

收稿日期：2023−11−02；修回日期：2023−12−27
基金项目：国家自然科学基金（62201576, U1833107）；中央高校基本科研业务费专项资金（3122022050）
作者简介：谢丽霞，博士，教授，主要从事网络与系统安全、信息安全方面的研究。
*通信作者 E-mail: yhyxlx@hotmail.com

列表的检测方法虽然简单易用，但对于未知的网络钓鱼反应较慢。基于启发式的方法采用一定的规则，但攻击者可以比较容易地规避这些规则。机器学习检测方法可能会受到特征规避的影响，而深度学习方法则可能面临可解释性差的问题。基于图的检测方法虽然展现出潜力，但构建图的过程比较烦琐。同时，攻击者也在尝试规避这些检测方案，使用基于浏览器和基于客户端的伪装技术[3-7]，甚至干扰网络安全爬虫，从而导致数据采集错误，造成检测误判[8-9]。

鉴于这些检测方案的局限性以及攻击者所带来的威胁，有必要深入研究网络钓鱼检测方案，以了解现有方法优缺点及存在的问题。为此，本文对网络钓鱼检测工作进行了调研。通过系统性地研究相关文献来尝试回答以下问题：网络钓鱼检测包括哪些检测场景？网络钓鱼检测方法有哪些？研究者使用的网络钓鱼检测数据主要为哪几类？使用不同的数据会给检测方法带来什么问题？常用的数据来源和数据集有哪些？各种网络钓鱼检测方法所面临的检测困境是什么？

基于以上问题，首先搜索了近 4 年的网络钓鱼检测相关工作，通过人工筛选出近 4 年的 82 篇高质量文献作为研究对象。从检测场景、数据集和检测方法等方面对提取的文献数据进行分析总结，强调了一些重要发现，最后总结了该领域存在的问题与挑战。

1 网络钓鱼发展及相关定义

1.1 网络钓鱼定义与发展

1.1.1 网络钓鱼定义

网络钓鱼旨在通过伪装成合法实体或组织，诱使受害者提供敏感信息、点击恶意链接、下载恶意文件或执行其他有害操作。这种攻击常常运用社会工程学、欺骗和伪装等手段，以实现窃取信息、盗取账号密码、传播恶意软件等恶意目标。作为威胁网络安全的一种重要因素，网络钓鱼由于其巧妙的伪装和欺骗性质，经常使用户在毫无察觉的情况下遭受损害。

1.1.2 网络钓鱼发展

网络钓鱼主要经历了 5 个发展阶段。

1）早期阶段：最早的网络钓鱼始于 1995 年，攻击者主要通过向用户发送虚假的电子邮件试图欺骗用户，攻击者采用相对简单的伪装，如拼写错误、语法错误等。但由于用户对网络钓鱼攻击的认识较低，导致用户很容易受到欺骗。

2）专业化发展阶段：攻击者使用更具有欺骗性的邮件和钓鱼网站。并利用一些社会工程学手段试图欺骗用户，如声称来自金融机构、使用威胁或紧迫性内容等。这个阶段的网络钓鱼攻击通常还会结合恶意软件，使攻击更具破坏力，同时该阶段攻击者开始有选择地攻击用户。

3）多渠道发展阶段：由于移动应用的快速发展，攻击者开始尝试在多种渠道进行攻击，如社交媒体、移动应用、短信等。攻击者通过伪装成合法应用程序、精心设计社交媒体帖子以及短信消息。

4）定制化攻击阶段：攻击者使用高级技术和加密方法，攻击者可能长期存在于受害者网络中，窃取敏感信息，通常与高级持续威胁（APT）攻击联系紧密。同时攻击者使用的社会工程学手段更有针对性和深度，涉及更多社会工程心理学原理[10-12]。

5）技术演进阶段：攻击者手法更加复杂，特别是在突发事件影响下，借助人工智能和机器学习等技术，进一步提升攻击的隐蔽性和伪装效果[13-14]。当前，常见的伪装技术包括基于客户端和服务端的策略。基于客户端的伪装主要通过在访问者浏览器中运行的 JavaScript 代码实现，利用 Cookie、鼠标移动轨迹等属性进行过滤；基于服务端的伪装则通过解析 HTTP 请求中的信息识别用户[3-7]。此外，攻击者还频繁采用中间人攻击[15] 以及针对检测模型的对抗性攻击[13] 等。

1.2 网络钓鱼场景

网络钓鱼检测主要集中在互联网媒介上，如网站、电子邮件、区块链、物联网和社交网络等场景中，但网络钓鱼的检测场景不仅局限于上述的 5 种类型。短信、电子传真、即时通信等媒介也被证实可以作为网络钓鱼检测的场景，这些媒介扩展了网络钓鱼的威胁范围[16]。图 1 呈现了网络钓鱼的媒介、传播场景以及攻击过程中所使用的技术手段之间的关联关系。这一综合的分析有助于研究人员更好地理解网络钓鱼的多样性和复杂性，从而为进一步的防御策略提供有力的支持。

在短消息媒介和语音媒介方面，隐私问题和数据安全等方面的担忧导致检测工作研究进展缓慢。可以考虑从如下 3 方面，确保研究工作的顺利进行。1）一些研究者提出的安全协议[17-18]、数据匿名化和加密等方法可以作为解决该问题的一种思路。2）在设计检测算法时，引入差分隐私技术和

联合训练技术，以确保数据不会被未经授权的人获取，从而增强隐私保护。3）为增强用户对算法操作的信任感，采取措施确保用户能够理解模型如何使用他们的数据及如何做出决策。

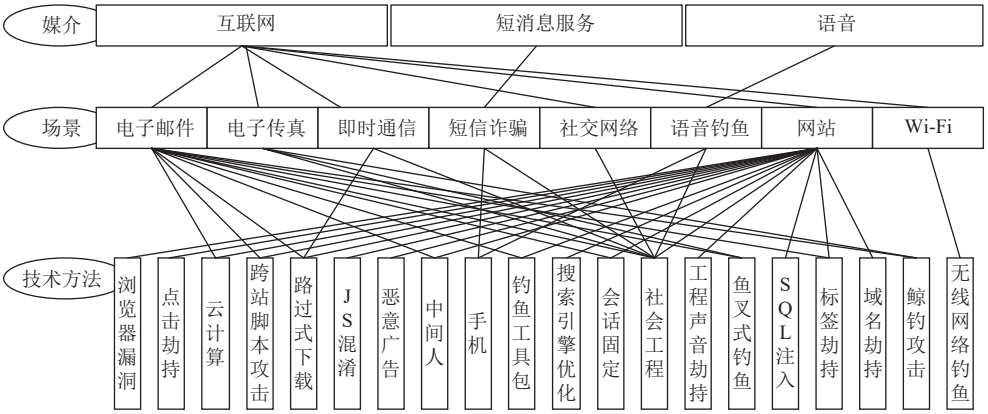

图 1　网络钓鱼媒介、场景以及攻击技术手段之间的关系

1.3　网络钓鱼检测阶段及检测方法

1.3.1　网络钓鱼检测阶段

网络钓鱼的攻击过程可以大致划分为 5 个阶段：目标选择和侦察阶段、伪装准备阶段、攻击向量生成阶段、传播阶段、受害者交互阶段和攻击完成阶段。但由于前 3 个阶段的黑盒性特点，研究者很难在这些阶段中检测出网络钓鱼行为。目前，对网络钓鱼的检测主要集中在传播阶段和受害者交互阶段。图 2 展示了网络钓鱼的攻击阶段、交互场景以及对应交互场景下所使用的数据情况。

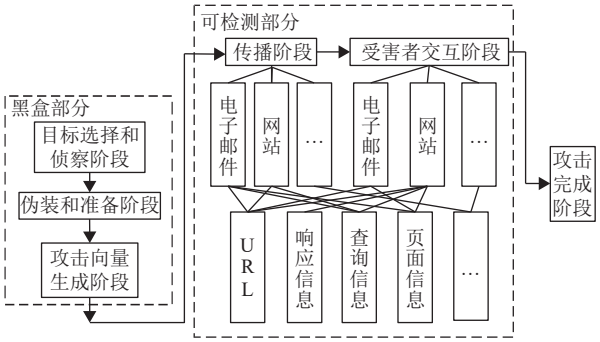

图 2　网络钓鱼攻击过程阶段与检测重点

1.3.2　网络钓鱼检测方法

根据文献中使用的具体的技术手段，对网络钓鱼检测方法及适用场景进行了如下的总结。

1）基于列表的检测方法

基于列表的检测方法主要依赖于构建黑白名单列表，其中包含恶意或可疑的 URL、IP 地址信息等。当用户尝试访问网站时，系统会将用户请求与列表进行匹配，然后根据匹配结果决定是否继续进行访问，以此来识别潜在的钓鱼网站。尽管基于列表的检测方法在某种程度上能够检测网络钓鱼，但在面对不断变化和复杂化的钓鱼技术时，其效果可能会受到限制。近期的研究表明，基于列表的检测方法主要以辅助的方式被整合到其他更先进的检测方法中，以提供更全面的保护[19]。

基于列表的检测方法在处理电子邮件、即时通信、短信诈骗、社交网络、网站等场景时表现出一定的效果。这些场景主要涉及文本信息和链接，通过建立包含恶意 URL、欺诈文本、特定关键词等内容的列表，可以进行安全防护。在 Wi-Fi 场景中，通过建立 AP 的指纹信息列表进行识别同样是一种有效的方法，包括加密方式、信道、ESSID、BSSID、网卡信息等。这种方法的优势在于其简单而高效，但存在漏报、错报等情况。

在电子传真场景中，由于其复杂的格式和多样性，通用的列表规则难以适应其特定的威胁模式。在语音钓鱼场景中，语音内容的动态性、多样性和关联性使得静态的列表无法有效跟踪最新威胁。

2）基于启发式的检测方法

启发式检测方法主要通过人工规则构建和特征提取，捕捉到网络钓鱼的一些常见特征和行为。在启发式检测方法中，人工规则的构建是关键步骤。研究人员需要深入了解钓鱼攻击的不同变种、攻击者的策略以及受攻击对象的行为，以构建能够准确识别网络钓鱼的规则。

启发式检测方法在电子邮件、社交网络、网站、Wi-Fi 等场景中表现出较好的适用性。主要通过分析识别异常模式和监测恶意特征，在一定程度上能够提供有效的安全防护。

然而，在一些特定场景如即时通信、电子传真和语音钓鱼中，启发式检测方法可能会面临一些挑

战，限制了其发挥作用的效果。在即时通信和电子传真场景中，由于缺乏明确的规范，传输的信息可能具有较大的变化性，使得常规的启发式规则无法准确捕捉到异常模式。而在语音钓鱼场景中，由于需要结合具体的语境进行分析，设计通用的规则变得更加困难。

3）基于机器学习的检测方法

基于机器学习的网络钓鱼检测主要是通过收集大量的数据，这些数据包括可疑的电子邮件、网页、URL、文本内容等进行标记、预处理等操作以构建数据集，然后，对数据进行特征工程以提取有用的特征。最后，通过选择合适的机器学习算法构建网络钓鱼检测模型，如逻辑回归、决策树（DT）、随机森林（RF）、支持向量机（SVM）、K 均值聚类（K-Means）等，并使用标记的数据进行有监督训练[20]。也有部分研究者使用无监督聚类的方式进行模型的训练以完成网络钓鱼检测[21]。

基于机器学习的检测方法在电子邮件场景中主要通过提取邮件的结构、文本和附件等特征进行识别。在即时通信、社交网络、电子传真和短信诈骗场景中，主要依赖于文本和链接等特征的提取，然而，由于数据的变化性，需要考虑如何处理可能缺失的特征，以提高模型的鲁棒性。在网站中，通过提取网站内容、图片、结构等特征进行识别。在 Wi-Fi 场景中，通过提取网卡信息、路由信息、ESSID、BSSID、加密方式、信道等特征进行识别。

在语音钓鱼场景中，尽管机器学习在语音识别领域取得了一些进展，但由于语音钓鱼涉及模拟真实人类语音、声音调节和语音特征伪装等技术，使得检测这类欺诈行为更为复杂。

4）基于深度学习的检测方法

基于深度学习的网络钓鱼检测方法与基于机器学习的检测方法相似，都需要对数据的收集、预处理、标记等操作。但与传统的机器学习方法不同的是，基于深度学习的检测方案无须进行特征工程，而是依赖于多层次的神经网络自动地从输入数据中学习特征，如卷积神经网络（CNN）、循环神经网络（RNN）、深度神经网络（DNN）、深度信念网络（DBN）、深度自编码器（DAE）等。

与机器学习检测方法的适用场景相似，但深度学习在语音钓鱼中可能具有较好的表现。其能够通过端到端的学习有效地捕捉语音信号的复杂时序结构和高级特征表示。

5）基于图的检测方法

基于图的网络钓鱼检测方法主要是通过将数据分为节点、节点属性、边以及边属性等，以反映数据各部分之间的关联。然后，利用图的结构和这些关系，进行数据的分析、预测和决策。网站、电子邮件检测场景中图的构建主要基于页面信息的相互联系、URL 之间的联系以及响应信息之间的联系等。即时通信检测场景中，构建用户之间的关系图，检测异常的通信模式和潜在的社交工程攻击。社交网络检测场景中，构建社交网络图，检测异常节点和连接，发现潜在的欺诈行为。区块链检测场景中，图的构建可以基于交易之间的相互信息以及地址之间的沟通信息。通过将区块链上的交易和地址映射到图的节点和边上，可分析和监测区块链网络中的信息传递和交易模式。

但基于图的方法可能在处理语音信号中的复杂模式和特征上缺乏足够的灵活性，因此不太适用于语音钓鱼检测。语音钓鱼涉及多样的声音特征和变化，而图结构可能难以有效捕捉这些复杂性。在电子传真场景中，电子传真通常以文本和图像形式进行通信，而通信内容的结构可能相对简单，不太适合采用图模型。在 Wi-Fi 检测场景中，Wi-Fi 通信涉及网卡信息、路由信息、ESSID、BSSID、加密方式、信道等底层特征，这些特征更适合通过其他方法进行提取和分析，而不一定能够充分利用图模型的优势。

表 1 总结了现有检测方法的基本特点以及适用场景，其中基于深度学习和基于机器学习的研究方法占大多数，且具有较好的准确率。

表 1　现有检测方法及适用场景

检测方法	基本特点	电子邮件	电子传真	即时通信	短信诈骗	社交网络	语音钓鱼	网站	Wi-Fi	相关文献
基于列表的检测方法	检测数据匹配列表内数据完成检测	√		√	√	√		√	√	[22-23]
启发式检测方法	检测数据通过设计的规则依据权重完成检测	√			√	√		√	√	[24-25]
机器学习检测方法	选择和提取特征选择合适的机器学习算法完成检测	√	√	√	√	√		√	√	[26-27]
深度学习检测方法	自动从数据中提取高维特征完成检测	√	√	√	√	√	√	√	√	[28-29]
基于图的检测方法	依据数据之间的关系完成检测	√		√	√	√		√		[30-31]

2　网络钓鱼检测数据、数据集分析

2.1　样本数据分析

网络钓鱼作为一种复杂的网络威胁手段，涉及多种媒介和攻击方式，对数据的多样性和丰富性提出较高要求。研究者在进行网络钓鱼检测时需考虑不同的媒介和数据来源，以全面捕捉潜在的威胁。但不同的媒介可能会产生不同的特征和模式，需要灵活的数据处理和分析方法。面对这样复杂的检测现状，研究者需要在设计检测算法和选择数据集时权衡不同的因素。有些研究者可能更注重快速地检测判断，此时可能会选择使用响应快或无须响应的数据，如仅使用 URL 进行特征采集[32]。而为了提高检测的可解释性并获得更全面的威胁情报，研究者可能会选取更多的数据，如 URL、页面信息以及页面截图等。同时，一些研究者还在不同的检测场景下探索使用新的检测数据[33]。

图 2 简要展示了当前检测方法所包含的主要样本数据（URL、页面信息、响应信息、查询信息），本节将对这些数据作详细分析。

2.1.1　URL 样本数据分析

URL 指用于在互联网上定位资源的字符串，通常由协议、域名、路径、查询等信息组成。URL 在网络钓鱼检测中的关键作用主要源于其域名滥用的情况[34]。攻击者利用域名的相似性，注册与合法网站域名类似的虚假域名，如将合法的“www.example.com”注册为“www.examp1e.com”。此外，攻击者在 URL 路径、查询中插入用户将要访问的 URL，如“wwww.abc.com/?href=www.example.com”，这种微小的差异会让用户疏忽，从而导致网络钓鱼攻击。此外，URL 的滥用不仅限于虚假域名和路径，也涉及缩短 URL 和重定向 URL 等[32]。基于 URL 的检测方法[29, 35-37]被研究者广泛使用于网络钓鱼检测中，这种方法无须加载网页内容，能够快速进行检测。然而，基于 URL 的网络钓鱼检测方法也有其限制，如无法有效地检测缩短 URL、重定向 URL 等[38-39]。

URL 网络钓鱼检测常用特征如图 3 所示，这些特征有助于识别潜在的网络钓鱼。但随着攻击者技术的不断演进，单纯依靠 URL 进行网络钓鱼检测并不能准确识别[28, 39]。

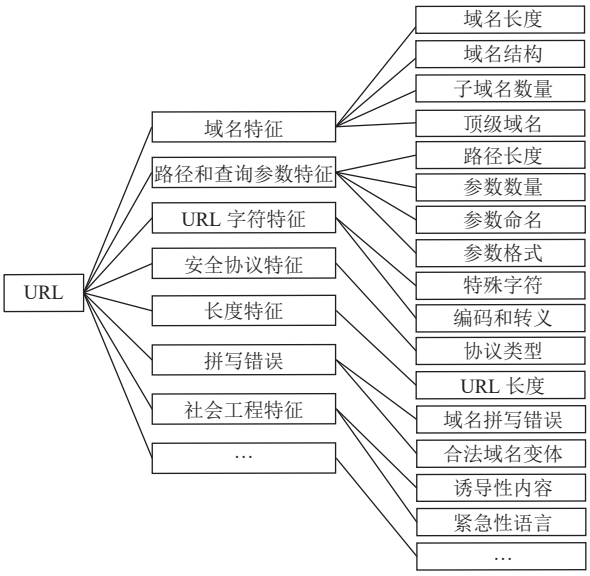

图 3　URL 常用特征

2.1.2　页面信息样本数据分析

页面信息指通过 URL 进行访问的网页信息，其中包括 HTML、文本、JS、CSS、图片等。获取阶段分为传播阶段和受害者交互阶段。传播阶段，攻击者依托合法网站进行传播，钓鱼信息可能仅是页面中的一个超链接。因此，页面信息的主要价值在于受害者交互阶段。一些研究者已经开始利用页面信息来提高网络钓鱼检测的准确性。如文献 [40] 通过对页面进行截图并提取组件布局信息，创建新的特征，从而增强检测能力。此外，文献 [28] 从页面信息中提取文本信息，以构建更具可解释性的模型。尽管从页面信息中提取特征可以获得更多的信息，但通常需要加载网页，这可能会在检测过程中引入一定的延迟[39]。因此，在设计网络钓鱼检测系统时，需要权衡检测准确性和响应速度之间的关系，以确保系统在实际应用中具有可接受的性能。页面信息特征提取时的常用特征如图 4 所示。

2.1.3　响应信息样本数据分析

响应信息指通过 URL 进行访问时服务器对客户端响应中的数据。与页面信息相似，有效的特征提取的重点会放在受害者交互阶段。但特征提取会十分复杂，需要考虑用户与恶意网站的互动。这些特征可能包括用户的点击行为、输入信息的处理方式、网页内容的变化等。文献 [7, 41-42] 尝试从响应信息中提取特征以进行网络钓鱼研究。文献 [42] 从 HTTP 请求头中收集特征，这些特征可以用作检测的依据。文献 [7] 通过分析 HTTP 请求中的配置项特征来研究攻击者的伪装行为。响应信息中的常见特征如图 5 所示。

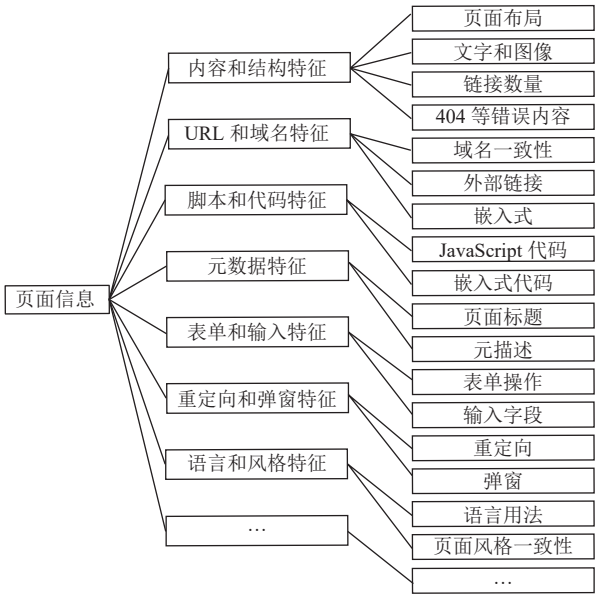

图 4 页面信息常用特征

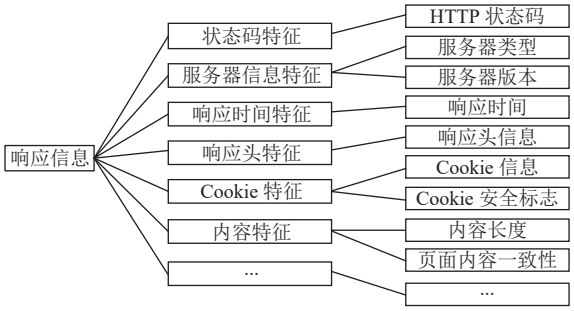

图 5 响应信息常用特征

2.1.4 查询信息样本数据分析

查询信息需要主动查询获取，查询信息能够显著提高检测速度，实现类似基于 URL 的检测方法的快速响应能力[33]。查询信息的利用可以提高检测速度和能力。但需要仔细考虑如何融合查询信息特征以提高检测准确性，同时应意识到查询信息不是始终可用的，因此需要综合考虑多种特征和方法来进行网络钓鱼检测。常见的查询信息特征如图 6 所示。

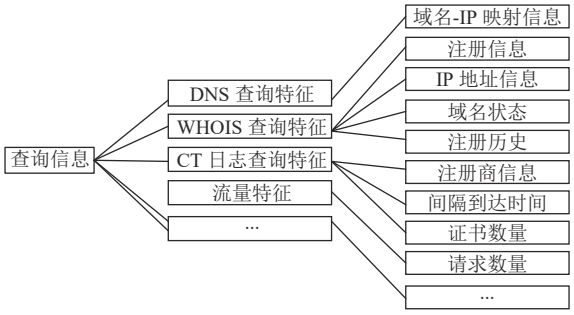

图 6 查询信息常用特征

2.1.5 分析总结

本小节对数据样本进行了分析，主要分析了 URL、页面信息、响应信息、查询信息。在网络钓鱼检测中这 4 类信息使用较多，除此之外还有区块链的交易信息在区块链中进行网络钓鱼检测，以及社交文本信息，在社交网络中进行网络钓鱼检测。表 2 对样本数据进行了优缺点总结，在实际应用过程中要根据需要选择合适的数据，如侧重实时性应选择无须访问的检测数据，侧重精度应选择多种数据进行综合评估检测。

表 2 网络钓鱼检测数据优缺点总结

样本数据	优点	缺点	相关文献
URL	检测速度快	对重定向和缩短URL识别较差	[29, 35]
页面信息	检测精度高	需要加载网页，无法访问时不能有效检测，且不能有效检测多语言钓鱼	[26, 40]
响应信息	检测精度高	检测速度慢，无响应时不能给出有效检测结果	[43]
查询信息	不易被规避	查询信息可能缺失，影响检测精度	[44]
交易信息	特定场景具有优势	数据处理烦琐	[31]
社交文本信息	特定场景具有优势	数据获取难，数据提取难	[45]

2.2 数据集分析

在分析数据集时发现如下问题。

1）多数研究者未公开自己实验时的数据集和实验的代码，造成了模型复现的困难。

2）收集和使用数据时研究者没有注重如下几点：数据的重复性（如使用 PhishTank 进行数据收集时，获取的数据中存在大量重复 URL）、数据的时效性（部分数据集的老旧数据特征分布与当前网络钓鱼数据特征分布不同）、数据的公平性（如合法 URL 无路径信息或 http://、https://仅存在于钓鱼 URL 中）、数据的正确性（如由于攻击者会采用伪装等技术阻止网络爬虫获取正确的访问信息）以及数据标记的准确性（数据标记不准确可能导致模型性能不能准确体现）。

3）研究者公开的数据集存在参差不齐的状况，如 ISCX-URL2016 数据集，由于包含大量重复数据，导致在使用数据进行模型训练时精度异常，部分研究者在该数据集上的精度接近 1，并不能反映出模型的性能。这要求在选择使用数据集时应对数据集进行鉴别，以反映出真实的模型性能。

在数据集的收集和使用方面，研究者需要考虑数据的重复性、时效性、正确性和准确性。网络钓鱼数据集的选择与使用对实验结果产生重要的影响。

在模型的性能评估方面，建议应在小型数据集（小于 10 万）、中型数据集（大于 10 万小于 100 万）和大型数据集（大于 100 万）上进行测试，以全面地对模型进行评估。

表 3 总结了研究者在网络钓鱼检测场景中的数据来源情况以及数据集使用情况。

表 3　网络钓鱼数据来源以及数据集

场景	名称	类型	数据分布情况	获取地址
网站、物联网	PhishTank	URL	钓鱼	www.phishtank.com
	Openphish	URL	钓鱼	www.openphish.com
	Alexa	URL	合法	www.alexa.com/topsites
	5000best	URL	合法	5000best.com/websites/
	ISCX-URL2016	URL	35 300:10 000	www.unb.ca/cic/datasets/url-2016.html
	Ebub2017	URL	36 400:37 175	github.com/ebubekirbbr/pdd/tree/master/
	PhishingDataset	URL	800 000:759 485	github.com/vonpower/PhishingDataset
	VisualPhish	页面截图	9 363:1 195	sahar.abdelnabi@cispa.saarland
电子邮件	MDP-2018	特征向量	5 000:5 000	data.mendeley.com/datasets/h3cgnj8hft/1
	EnronEmailDataset	电子邮件	合法	www.cs.cmu.edu/～./enron/
	JosePhishingDataset	电子邮件	钓鱼	monkey.org/～jose/phishing/
	Millersmiles	电子邮件	钓鱼	www.millersmiles.co.uk
	IWSPA-AP 2018	电子邮件	36 400:37 175	github.com/ebubekirbbr/pdd/tree/master/
区块链	infura	交易数据	合法	www.infura.io
	etherscan	交易数据	钓鱼	etherscan.io

3　钓鱼检测方法分析

3.1　基于列表的检测方法

3.1.1　检测方法研究

基于列表的检测方法主要是与其他检测方法融合进行检测。文献 [22] 融合传统白名单检测方法和页面的视觉相似性，在 6 个不平衡数据集上平均精度达到 96.17%。文献 [19, 23] 通过融合传统列表方法、启发式方法、机器学习方法和页面的视觉相似性分别提出了两种网络钓鱼检测方法。文献 [19] 的方法具有较高的准确率和实时性；文献 [23] 的方法则需要较高的响应时间。

3.1.2　黑名单更新方式研究

基于列表的检测方法面临的主要难题在于列表的创建和维护，通常可以将创建和维护方法分为被动方式和主动方式两种。被动方式主要包括以下几种途径：来自安全厂商的信息、用户的举报、第三方威胁情报、恶意数据库的数据，以及在访问时进行的规则判定。主动方式则包括采用域名生成技术的定向网络爬取[6]、结合种子列表的拓扑网络爬取，以及全网爬取等方法对信息抓取并进行判断，用于创建和更新检测列表。

3.1.3　有效性研究

基于列表的检测方法虽然能够快速过滤出列表中的网站，但却无法有效抵御网络钓鱼的零日攻击。因为钓鱼网站的生命周期短暂，两种列表构建与更新方式难以在短时间内捕捉到这些新出现的钓鱼网站。此外，攻击者会根据访问者的请求信息进行判别，正常用户返回钓鱼页面，网络安全爬虫返回正常页面，从而干扰检测方法的有效性。一些研究者对此进行深入研究，如文献 [8-9] 评估爬虫对规避技术的反应。发现黑名单无法检测出一类基于 JavaScript 的规避行为，并且攻击者可以利用伪装向量进行攻击，从而对黑名单的更新和扩展产生影响。而文献 [4] 则通过对 2 883 个热门网站进行错字抢注和组合抢注，发现黑名单方法对于伪装网站不能起到有效作用。

3.1.4　基于列表的检测方法面临的困境

基于列表的检测方法对于未识别的网络钓鱼识别缓慢，且维护一个实时、准确的列表需要大量的时间和资源。而网络钓鱼攻击者经常更新、更改钓鱼信息和规避检测，因此要保持列表的准确性需要不断地更新和验证。同时，基于列表的网络钓鱼检测方法通常是一种黑盒方法，用户无法理解系统是如何判断一个网站是否是钓鱼网站，这可能导致用户对系统的结果产生不信任。基于列表的网络钓鱼检测方法的对比分析总结如表 4 所示。

表 4 基于列表的网络钓鱼检测方法总结

文献	检测方法	使用数据	优点	缺点
文献[22]	视觉相似性（白名单）	URL，页面信息	模型泛化能力相对较强	使用的数据集小，不能反应模型性能
文献[19]	黑名单，视觉相似性，启发式方法，机器学习	URL，页面信息，响应信息，查询信息	设计3 000个特征，准确率达到99%，实时性强	特征提取烦琐，数据泛化能力不强
文献[23]	黑名单，视觉相似性，启发式方法，机器学习	URL	检验基于启发式和基于黑名单的过滤器的重要性	系统需要较长响应时间

3.2 基于启发式的检测方法

3.2.1 检测方法研究

启发式检测方法通过构建规则进行网络钓鱼检测，文献 [24] 基于 URL 信息和页面信息中的 30 个特征，为特征分配一定权重，构建了一种规则树的分类模型。而文献 [25] 则对 URL 和页面信息进行 Google 搜索和必应文本识别，通过匹配查询结果做最终检测判定。两种方法都需要进行访问页面，相较于文献 [24] 提出的方法，文献 [25] 由于引入外部的查询模型的实时性会极大地受到查询结果的影响，同时 Google 搜索和必应文本识别，在模型中所占的权重极大，若无法给出有效反馈可能导致模型精度的下降。

3.2.2 启发式检测方法面临的困境

由于钓鱼攻击的不断演变和多样性，人工规则构建可能会显得不够灵活且适应性较差。攻击者可以通过微小的变化规避这些规则，从而降低检测方法的有效性。虽然可以从 URL、页面信息、响应信息和查询信息中提取各种特征，但钓鱼攻击中使用的技术和方法在不断变化。因此，仅仅依赖已知的特征可能无法覆盖所有的钓鱼形式。

同时，设计的启发式规则可能过度关注某些特征或模式，而忽略其他特征或模式，这可能导致启发式方法在实际应用中仅在特定数据集上表现良好，而对于其他情况效果不佳。攻击者通过了解规则的工作原理，采取措施隐藏或模糊网络钓鱼的特征。因此在面对未知的、全新的钓鱼攻击时，效果较弱，这意味着对抗零日攻击可能会有限。基于启发式检测方法的对比分析总结如表 5 所示。

表 5 基于启发式网络钓鱼检测方法总结

文献	检测方法	使用数据	优点	缺点
文献[24]	规则树	URL，页面信息，查询信息	可解释性强，计算资源消耗低	存在特征规避的风险，特征提取存在缺失可能
文献[25]	组合规则	URL，页面信息，查询信息	视觉和文本身份的混合化	存在逻辑漏洞，模型不具有实时性

3.3 基于机器学习的检测方法

机器学习领域的许多算法已经被广泛应用于网络钓鱼检测，如随机森林（RF）、支持向量机（SVM）、极限学习机（ELM）、K 均值聚类（K-MEANS）、逻辑回归等。依据研究者的研究方式的不同，将基于机器学习检测方法划分为单一分类器研究和多种分类器研究以及检测分析性研究。

3.3.1 单一分类器研究

早先的研究者通过多级架构、数据降维、特征选择等方法以提高模型效率和准确性。

1）多级架构、数据降维

多级结构主要通过组合不同的检测方法，利用这些检测方法的优点，实现对网络钓鱼的精确检测。文献 [26] 通过白名单和防伪标识进行快速过滤，然后通过多尺度 CASE 特征检测模型进行精确识别，而数据降维主要是为了降低计算负担，加快模型的推理与训练。文献 [46] 则通过降噪自编码器（SDAE）降低数据维度，从 URL 和页面信息中提取表面特征、拓扑特征以及深度特征。文献 [26] 所提方法具有一定的实用性，但数据集中包含的 URL 不够丰富。而文献 [46] 所提方法仅在平衡数据集上有效。

2）特征选择

特征选择的目的主要是为了减少冗余特征，降低计算成本，使相关算法能更高效、快速地完成分类。常见的特征选择方法主要有过滤法（独立于机器学习算法进行特征评估和排序）、包装法（使用特定的机器学习算法来评估特征的贡献）和嵌入法（结合特征选择和模型训练，通过机器学习算法自动选择特征）。文献 [47-48] 采用过滤法进行特征选择，文献 [47] 使用信息增益（IG）、CS（Chi-Square）和 RfF（Relief-F）等测量方法对从 URL 中提取的 46 个原始特征进行选择，保留最重要的 9 个特征。而文献 [48] 则通过互信息方法识别两个类别之间的冗余特征。文献 [49-50] 采用包装法进行特征选择，文献 [49] 从 URL、页面信息和查询信息中提取特征，通过前向选择确定最优特

征数量。文献 [50] 从 URL、响应信息和查询信息中提取特征，使用递归消除进行特征选择。文献 [51] 则通过嵌入法完成特征的选择，通过 MOE 组件进行特征选择，在准确率和召回率方面优于 MOE-SVM 等其他模型。

特征选择一定程度上能提升模型的性能，但特征选择容易忽略特征之间的相互关系，并且随着时间的变化，数据特征可能发生改变造成之前选择的特征不能适应新的数据分布，因此在特征选择前应考虑数据是否具有这种相关性和变化性。

3.3.2　多种分类方法研究

研究者近期主要通过评估不同分类器来确定最合适的分类器，通过评估精确度、召回率、F1 得分等，确保选择的分类器在解决问题时能够取得最优效果。

文献 [43] 从 URL 和页面信息中提取 4 组特征，前 3 组特征由文献 [52] 提出。通过比较不同分类器的性能，最终确定在 LightGBM 上性能最优。文献 [53] 在 URL 和页面信息的基础上增加对响应信息和查询信息的特征提取，确定使用 RF 具有较好的分类效果。而文献 [27] 使用 ImageNet 数据集训练深度学习图像分类模型，使用 VGG16 从网站截图中提取特征，评估发现 LR 进行分类具有较好的准确性。文献 [54] 对 URL 进行特征的提取，选取 9 种词汇特征，其中在 RF 上的检测精度达到 99.57%。

由于网络钓鱼检测是一个二分类问题，将样本依据特征划分为不同的组或簇，也可实现对网络钓鱼的检测。如文献 [21] 提取 7 类电子邮件特征，将 7 类特征输入 K-means、DBSCAN 和 Hierarchical Agglomerative 这 3 种聚类算法中进行分析，发现 K-means 算法在性能上优于其他聚类方法。

3.3.3　检测分析性研究

为了应对对抗性攻击，文献 [55] 通过向训练数据中添加或插入噪声，降低分类器内特征的权重，增强其他特征，进而构建 N 元分类系统，测试发现所提出的模型相对于原生模型更具鲁棒性。在探索特征之间相关性方面，文献 [56] 对 URL 的静态特征组合进行定性分析，寻找特征之间可能存在的关系。为了评估不同机器学习算法的性能，文献 [57] 在公开数据集上对经典分类算法进行评估，发现多层感知机、决策树和集成型算法（如随机森林、梯度树提升和 AdaBoost）表现最优，而 SVM、K-最近邻算法和朴素贝叶斯性能较差。

3.3.4　机器学习检测方法面临的困境

基于机器学习的网络钓鱼检测方法在捕捉复杂模式和适应未知攻击方面具有显著的优势。然而，也面临着一系列挑战，这些挑战包括数据质量、对抗攻击和概念漂移等。

1）数据质量方面。主要来源于两个方面。一方面是数据获取问题，研究者在进行钓鱼网站检测研究时需要获取各种数据，如 URL、网页信息、响应信息、查询信息等。然而，攻击者会采取伪装和规避等措施[3-7]，阻止研究者获取正确的信息。如攻击者故意隐藏或修改钓鱼网站的特征，使其难以被准确地采集。另一方面是数据标记问题。确定一个网站是否为钓鱼网站通常需要根据一些标准或投票方式进行判定。然而由于错误标记，难以确定网站的真实性质。此外，有些研究者尝试使用小范围的数据标记，然后使用聚类的方法完成数据标记[47]，但所选特征不一定能够完全反映数据的真实分布，从而导致错误的标记结果。

2）对抗性攻击方面。攻击者采取多种策略来欺骗机器学习模型，这些策略包括特征空间进行扰动[13]、规避重要特征、使用 GAN 生成对抗性样本进行对抗性训练[58-59] 等，这些策略都可以用来针对模型进行攻击。

3）概念漂移方面。由于网络钓鱼攻击技术不断变化，导致数据的特征分布发生改变，使原有模型无法很好地捕捉新的数据特征，从而导致检测精度下降。少数研究者注意到这方面的问题并给出了一些解决方法[60]。

基于机器学习的网络钓鱼检测方法的对比分析总结如表 6 所示。总体而言，基于机器学习的网络钓鱼检测更多的是从数据提取范围、特征选择、对抗规避、多种方法评估等方面进行研究。少数研究通过组合设计改进检测模型。

表 6　基于机器学习的网络钓鱼检测方法总结

文献	检测方法	使用数据	优点	缺点
文献[26]	多阶段过滤，RF	URL，页面信息	使用多阶段检测方案，可快速过滤合法网站	数据集样本类别不够丰富
文献[46]	ELM	URL，页面信息	解决样本不平衡问题，降低数据维度	特征提取耗时，特征缺失影响精度，存在特征规避

续表

文献	检测方法	使用数据	优点	缺点
文献[49]	逻辑回归	URL，页面信息，查询信息	对特征排序，遍历添加的特征进行选择	特征易被规避，不具有实时性
文献[50]	递归消除，RF	URL，页面信息，响应信息	递归消除降低特征数量，云上部署模型	不具有实时性，未考虑云上部署负载
文献[47]	RF	URL	使用信息增益等方法对特征进行消除	特征易被规避，无法检测缩短URL和重定向URL
文献[51]	MOE，RF	电子邮件特征向量	具有高召回率和高准确率的特点	在真实应用场景中可能实时性不高
文献[48]	KNN，ANN	电子邮件页面信息	设计堆叠分类器和软投票分类器	数据集小，仅在不平衡数据集上进行测试，实时性不强
文献[43]	多种分类方法评估	URL，页面信息，响应信息	提出27个针对网络钓鱼检测的新特征	特征存在缺失可能并影响精度，计算资源消耗大
文献[53]	多种分类方法评估	URL，页面信息，响应信息，查询信息	在8种机器学习和3种深度学习算法上进行评估	依赖第三方查询数据
文献[27]	多种分类方法评估	页面信息	11种机器学习和5种深度学习算法上进行评估	在不平衡数据集上效果不佳
文献[54]	多种分类方法评估	URL	提取9种词汇特征	未在多种数据集上评估，选取的特征简单
文献[56]	多种聚类方法评估	电子邮件页面信息	在3种聚类方法上对设计的7类特征进行分析	检测精度低

3.4 基于深度学习的检测方法

深度学习作为机器学习的分支领域，与传统机器学习最显著的不同在于其利用多层次的神经网络模型来自动地从数据中学习特征表示。依据使用数据的不同将深度学习检测划分为基于 URL 的深度学习检测方法、基于 URL 和页面信息的检测方法和其他检测方法。

3.4.1 基于 URL 的深度学习检测方法

基于 URL 的检测模型主要是从 URL 序列中提取字符和序列特征。文献 [60] 第一批使用 CNN 从 URL 中提取特征进行钓鱼网站检测。为了提取到多层次的特征，文献 [61] 通过不同大小核的单层 CNN 进行多层次特征提取，考虑到对 CNN 进行激活可能导致特征间变得强相关，且网络钓鱼 URL 和良性 URL 具有很强的空间相关性，引入 SpatialDropout1D 而不是标准 dropout 进行激活以增强模型泛化性能，在 5 个公开的数据集上取得 99% 的检测精度。仅从 URL 的字符特征出发，可能会忽略掉某些特征形式，文献 [62] 综合考虑 URL 的字符级和单词级嵌入表示，采用门控神经网络对两种嵌入进行合并整合，以防止直接拼接导致字符级对单词级的淹没问题。其采用多尺度 CNN，提取不同的特征表示，与文献 [61] 不同的是增加了深度金字塔模块对全局以及更高层次的语义提取，比单一特征信息的方法表现出更好的性能。部分研究者将研究的重点放在模型的功耗及推理时间上，文献 [63] 尝试采用 LSTM 和 GRU 构建两种模型，通过使用更少的参数，并在此基础上提升模型能，发现采用 GRU 具有较高的检测精度和较少的推理时间。同时，所提出的模型能在移动设备和 Raspberry Pi-4 上使用，并有较高的检测精度，但推理时间明显高于非移动设备。

注意力机制能使模型关注数据中的关键信息，通过在网络的不同层次中增加该机制，增强模型的灵活性和可解释性。文献 [35] 对每个 URL 提取字符级以及 3 个不同的 *n*-gram 级别表示，然后分别利用 CNN 提取分层特征，BiLSTM 提取时序特征，最后模型通过注意力机制对循环层的隐藏状态进行加权计算，确定各状态对应的权重，在自建数据集上的检测精度达到了 99.27%。但对该模型的源码运行测试发现，模型检测所消耗的时间大部分集中在 3 个不同 *n*-gram 的提取上，如何优化这部分的时间消耗是值得考虑的问题。文献 [64] 提出的 CNN-MHSA 模型，利用 MHSA 捕捉 URL 字符串之间的内在联系。具体地，将 URL 嵌入矩阵复制为两个副本，一个副本通过 CNN 提取 URL 的特征表示，而另一个副本则利用多头自注意力（MHSA）机制计算特征的权重，最后利用 MHSA 和残次链接完成对两个副本的组合以得到最终结果。但文章使用的自建数据集，通过对数据来源进行分析，钓鱼和合法数据之间存在明显的区分特征，作者并未考虑对该部分进行可信性和公平性处理。文献 [65] 先通过 Transformer 对 URL 进行编码再通过卷积对 URL 嵌入矩阵进行初步特征提取，并复制提取后的特征，然后采用两个并行模块（Transformer 模块、卷积模块）分别捕捉 URL 的全局特征、位置信息、内部依赖关系以及局部相关性，最后利用 Transformer 对两模块进行特征融合。模型在多

处使用了 Transformer，参数量较大，可能导致在资源受限的环境中训练和部署的问题，能否对编码部分以及融合部分进行替代，以加快模型训练及减少检测时间值得考虑。

URL 的统计特征主要是通过对 URL 中特殊字符、长度、是否包含特定单词等进行特征提取。为提取更全面的特征，文献 [29] 使用时间卷积和因子分解机，分别学习 URL 结构特征和统计特征，最后通过自定进度学习融合统计特征和结构特征，在大型数据集上取得 99.02% 的准确率。但通过对文中使用的统计特征的研究发现，主要通过对特征设置一定的阈值使得特征的取值仅包含 0 或 1，该做法可能导致一些特征在提取时出现错误分类问题，进而影响整体检测效果。

SDN 控制器主要是将网络控制平面和数据平面分离，从而实现网络的集中控制和编程，降低客户端的处理压力。文献 [66] 将 URL 检测工作从用户个人客户端移至 SDN 控制器，模型首先对 URL 的一些特征进行提取，然后利用 RFE-SVM 对提取的特征进行选择，最后利用在控制器内的 FS-CNN 进行检测分类。但在 SDN 上进行检测还应考虑 SDN 所带来的检测延迟，单点故障导致整个网络的 URL 检测功能可能受到影响等问题，同时，预先对特征进行了提取以及选择，并不能完全发挥 FS-CNN 的作用。

基于 URL 进行网络钓鱼识别，相较于基于其他数据的检测方法具有快速响应的能力，但由于 URL 提供的信息相对有限，有时仅通过 URL 并不能准确识别网络钓鱼，需要结合具体的应用环境选择性地结合其他的一些信息如页面信息、响应信息等做进一步识别。

3.4.2　基于 URL、页面信息和查询信息的检测方法

文献 [40, 67-68] 是基于相似性的方法。文献 [67] 主要通过网页截图和 Logo 列表进行网络钓鱼识别，对于网页截图使用 Faster-RCNN 进行 Logo 的识别，然后通过 Siamese 模型对识别到的 Logo 与列表内的 Logo 进行相似度分析，若有匹配结果则分析 Logo 列表中对应的域名与检测的 URL 之间的差异进而完成检测。但该方法可能面临 3 点问题：对于不在 Logo 列表中的待检测目标识别差；对于不包含 Logo 信息的网站、具有多种 Logo 信息的钓鱼网站识别能力较差；仅通过捕捉 Logo 并不能完全代表页面语义信息。文献 [68] 则使用三元组卷积神经学习合法网站内部网页之间的相似性，若与合法页面不相似则归类为合法页面。但该方法仅适用于检测已知品牌的网络钓鱼页面，同时由于已知品牌可能为了传达出不同的品牌意图而造成假阳性或假阴性。文献 [40] 从动态分析和静态分析两个角度进行网页意图分析，静态分析通过使用 Faster-RCNN 提取 Logo 信息和可能的登录凭据页面入口信息，通过与网页交互递归寻找用户凭据页面并进一步进行静态和动态分析，直到找到登录凭据页面或达到最大访问层，最后通过与列表内的可信徽标进行相似度分析以进行检测。该模型存在一个潜在的问题是由于动态寻找登录凭据页面导致检测时间的不固定，导致用户对该模型检测有较大的心理波动。

针对电子邮件检测场景，文献 [69] 提出一种在无标记的数据上构建数据集的方法。通过对数据进行小范围手工标记，然后采用 KNN 和 K-Means 相结合的方法将与手工标记相似的数据聚类在一起，提出一种数据集构建方法，最后使用聚类的数据训练 LSTM 模型完成分类。但通过聚类的方法完成数据集的生成有一定错误归类的风险，造成后续检测精度异常。

为了增强模型的可解释性以方便网络安全专家进行分析性研究，文献 [28] 利用 LSTM 从 URL 和页面信息模态中分别提取字符特征和单词特征并利用 ResNet-50 从网页截图模态中提取特征，同时为每种模态添加注意力机制，以计算每个模态中分配给特征的重要性分数。然后采用共享字典学习方法用于在注意力机制中调整不同模态数据的表达。最后利用模态注意力，确定每个模态的注意力得分。模型在元素级别和模态级别具有很强的解释性，但模型发挥全部性能需建立在网站能正常访问的前提下，在缺少网页信息和页面截图的情况下模型精度下降巨大，能否在设计模型时考虑这种数据缺失，并尝试使用注意力机制等方法去调整这种缺失问题有待解决。

3.4.3　其他

针对区块链检测场景，文献 [70] 利用 BP 神经网络提取网络处理转移特征和状态特征，同时使用全卷积神经网络（FCN）和 LSTM 网络处理交易特征，然后通过拼接以进行全面表征。但模型采用先进行手工特征提取，然后再通过上述模型进行检测，这种方式并未发挥出神经网络的特征提取能力，导致模型检测精度并不高。针对社交网络检测

场景，文献 [45] 通过对聊天文本数据进行收集，使用 CNN 进行特征的提取，依据所提取特征信息将攻击划分出不同的攻击阶段，并提出一种基于有限状态机的攻击阶段检测模型，依据不同的攻击阶段与受害者进行互动提醒，并转移攻击阶段状态。但检测模型仍面临如下一些问题：实验数据小；仅能进行单句话检测，缺少上下文语义；仅能进行文字类检测；用户隐私问题等。

为了提高模型的泛化能力、减少过拟合、丰富数据样本以及平衡数据集，文献 [58-59] 使用 GAN 生成对抗性数据。文献 [58] 将 URL 转化为图片，使用 GAN 生成同形文字图像，但模型生成图像内容仅支持英文。而文献 [59] 则通过在 GAN 的生成器中使用 LSTM，并使用 CNN 鉴别器，构建一种仅需使用文本和嵌入，便能进行候选域名生产的方法。

3.4.4 深度学习检测方法面临的困境

基于深度学习的网络钓鱼检测方法相较于传统机器学习具有多方面的优势，如特征学习更充分、层次化抽象特征表示和端到端的学习等优点，但仍存在如下不足。

1）可解释性方面。由于处理的是高维度数据，数据分布和特征之间的关系更加复杂，导致深度学习检测方法的可解释性不如传统机器学习方法。

2）计算资源消耗方面。由于网络钓鱼具有多种检测场景，而深度学习模型通常需要大量的计算资源进行训练和推理，这对于一些资源受限的设备和环境来说可能是一个挑战。如文献 [63] 虽然可以运行在移动设备上，但推理时间可能无法被用户忍受。

3）数据选择方面。多数研究者仅通过对 URL 提取特征完成网络钓鱼的检测，但仅从 URL 提取特征可能无法检测一些类型的网络钓鱼，需要从更多的数据来源进行特征提取，但同时也要注意实时性、特征缺失等问题。部分研究者从网页截图中识别特征设计模型进行相似度分析，该方案应对已知的品牌具有较好的识别能力，但对于未知的品牌识别较差，从而导致模型泛化能力较弱。

同时也面临着数据质量方面、对抗攻击方面和概念漂移方面挑战。深度学习的网络钓鱼检测方法的对比分析如表 7 所示。

表 7 基于深度学习的网络钓鱼检测方法

文献	检测方法	使用数据	优点	缺点
文献[60]	CNN	URL	引入更新机制，可在移动设备上使用	模型泛化性不高，检测短URL和重定向URL具有一定难度
文献[62]	HDP-CNN	URL	解决字符特征对单词特征淹没问题	仅在不平衡数据集上完成测试
文献[63]	LSTM，GRU	URL	保证检测效率的同时具有较短的推理时间	检测短URL和重定向URL有一定难度
文献[35]	CNN，BiLSTM，注意力机制	URL	提出一个新颖的URL数据集	训练耗时长，检测反应慢
文献[64]	CNN，多头自注意机制	URL	利用多头自注意学习不同的特征权重	数据集存在大量合法域名无后缀，钓鱼域名有后缀
文献[65]	CNN，Transformer	URL	使用自注意机制和卷积模块捕获特征	新方法对比不充分，检测短URL和重定向URL有一定难度
文献[29]	时间卷积，因子分解机	URL	通过二分支捕获不同特征并进行分之间融合提取	检测短URL和重定向URL有一定难度
文献[66]	FS-CNN，SVM	URL	引入SDN将控制与数据分离，并进行特征选择	未考虑SDN负载情况，检测短URL和重定向URL有一定难度
文献[67]	CNN	URL，页面图像信息	提供一种可解释性方法，提出一种视觉检测方法	检测隐藏或没有Logo的页面有一定难度
文献[68]	CNN	页面图像信息	可对抗零日攻击，公开一个9 363张截图数据集	实时性差，对可信名单以外的检测不能做到精准识别
文献[40]	ResNET，CNN	URL，页面图像信息	抽象网页布局，并与登录凭证页面结合进行识别	实时性差，只能检测已知品牌
文献[69]	CNN，SVM，KNN，K-Means	URL，页面信息，查询信息	对特征排序，并遍历特征进行特征选择	特征易规避，不具有实时性
文献[28]	LSTM，ResNET，注意力机制	URL，页面信息，页面图像信息	从3个角度的进行可解释性设计，检测精度高	不具有实时性，计算资源消耗大
文献[45]	CNN	社交文本信息	有限状态机模型进行攻击相位的检测	数据集小，仅能对单句话分析，不能检测图片和语音
文献[70]	FCN，LSTM，BP	交易记录	BP神经网络处理转移特征和状态特征	仅针对以太坊平台

3.5　基于图的检测方法

基于图的检测方法主要根据不同场景下所使用的检测数据进行检测研究。

3.5.1　不同场景下基于图的检测方法

1）网站场景

网站检测场景研究者依据不同的数据构造图。如文献 [71] 基于被动 DNS 数据的域提出一种关联检测方案。首先通过提取特征将域名和 IP 划分为公用和专用。然后构建 4 种图（G-Baseline、G-IP、G-Domain、G-IP-Domain），最后对这 4 种图分别应用路径推理算法和 BP 神经网络算法进行图推理以完成检测，发现 G-IP-Domain 在检测方面具有较好的表现。

文献 [30] 对页面进行两轮邻接访问，进而对其中的整体超链接和网络结构进行图建模，通过对图进行分析提出了 17 个用于检测网络钓鱼的图特征，最后通过对比分析确定使用 C4.5 作为分类器具有较好的性能。但由于进行邻接访问，可能导致检测具有较高的时延，影响用户使用。

文献 [44] 则构建由 URL、域名、IP 地址、名称服务器和 URL 分割单词组成的异构图。考虑到一个顶点如果有大量良性邻居将被分类为良性这个问题，提出了一种边缘潜在分配机制。该机制通过计算不同实体之间的多种相似性和改进的兼容性矩阵，以及利用铰链损耗来分配边缘值。最后使用 LBP 进行推理。该方法比其他方法具有较高的 F1 得分。

2）电子邮件检测场景

文献 [72] 将图卷积网络（Graph Convolutional Network, GCN）应用于检测中。首先对电子邮件进行分词，然后将单词和电子邮件作为图的节点，边则表示单词的频率和单词的电子邮件频率。进而将文本分类问题转化为节点分类问题。最后利用 GCN 网络学习节点之间的关系。但模型所采用的数据集过于陈旧，主要以文字为主，数据特征分布与当前电子邮件有一定差异。

3）区块链检测场景

由于区块链具有分布式和链接性质，因此非常适合用于构建图结构，进行网络钓鱼检测。文献 [31] 根据交易的发送方和接收方将事务图分为发送图和接收图，然后对这两种图进行卷积操作以进行特征学习，同时结合交易双方的边缘特征性质，其中的一些特征转移至节点特征以进行再次卷积，最后将发送图和接收图提取的特征合并以进行最终分类。文献 [73] 则将网络钓鱼账户的检测建模为节点分类问题，应用随机游走对事务网络进行采样，从获得的子图中提取特征，利用 GCN 中进行嵌入学习，最后利用 LightGBM 中进行节点分类。文献 [31, 73] 共性问题是仅在以太坊数据集上进行测试，且检测精度相对较低实用性较差。

3.5.2　基于图的检测方法面临的困境

基于图的网络钓鱼检测方法面临一些重要的问题和挑战。

1）结构的动态性。由于网络钓鱼攻击者可能会频繁更改其攻击模式和基础设施，导致网络结构的动态变化。这使得基于图的检测方法在捕捉和适应这种动态性方面面临挑战，需要不断更新模型。

2）节点标签的噪声。节点的标签通常表示节点的属性信息，但这些标签可能受到误导或伪造。网络钓鱼攻击者可能会采取措施来模糊节点属性，使得基于节点标签的检测方法受到干扰。

3）图的构建选择。网络钓鱼数据有不同的来源，使得构建图的节点和边变得复杂。不同的构建可能会带来不同的检测性能，但图的构建方法并不总是明显。

4）计算复杂度。随着图规模的增大，图计算可能需要大量的计算资源和时间，限制方法的可扩展性。

表 8 基于图的网络钓鱼检测方法总结了对比分析的结果。总体而言，研究者更注重对图的构建进行研究，根据不同的场景数据创建出相应的图。然而，这些研究并未充分考虑实时性、检测精度以及检测方法的泛化性等方面的问题。

表 8　基于图的网络钓鱼检测方法总结

文献	检测方法	使用数据	优点	缺点
文献[71]	路径推理，BP	查询信息	基于被动DNS数据设计了域的关联方案	RF分类器的结果影响后阶段的推理
文献[30]	C4.5	URL，页面信息	构建网页及链接之间的图关系，并进行特征提取	存在一定的延迟
文献[44]	LBP	URL，查询信息	对URL进行信息的拆解，并完成图的构建	构建过程烦琐

续表

文献	检测方法	使用数据	优点	缺点
文献[72]	GCN	电子邮件页面信息	设计电子邮件的关联关系图	仅支持英文文本检测
文献[31]	CT-GCN	交易信息	根据交易信息的不同构建发送图和接收图	检测精度低，仅在以太坊上进行测试
文献[73]	GCN	交易信息	采用编码解码的学习过程	实用性差，仅在以太坊上进行测试

4 挑战及展望

4.1 面临的挑战

从攻击角度，如攻击者获取和复制用户的浏览器指纹，以欺骗基于风险的身份验证机制并消除双因素身份验证[74]，会给网络钓鱼检测带来新的难题。合法实体本身也存在漏洞挑战[75-76]。

从检测场景角度，部分特定场景的数据集难以获取，需要解决隐私问题、数据安全问题等。

从模型角度和数据角度，不同的检测方法和数据有着各自的优缺点。要提供更为全面的网络钓鱼防御机制，要仔细研究这些数据和模型的优劣。同时，目前大多数研究者采用的数据大多数为英文，尚未有研究者针对不同的语言进行检测研究。

从用户角度，用户可能存在对检测结果的偏见[10]、对钓鱼的成本认知存在差异[11]、对网络钓鱼的认知与辨别能力不足以及对钓鱼事件的语境反应[12] 等方面的挑战。

4.2 未来研究展望

4.2.1 可解释性

如何提供全面的可解释性，使用户能够理解模型如何根据输入数据做出预测或判断的能力[77]。一方面有助于快速改进和优化模型，另一方面，解释性信息对于开发教育和培训应用、游戏程序等方面也非常有用[78-82]，同时也能在设计网络钓鱼检测报告等方面发挥作用[83]，这有助于增强人们对模型的信任，提高对网络钓鱼的识别能力。

4.2.2 新特征的挖掘

由于网络钓鱼攻击的生命周期内涉及的数据范围广，如何挖掘出新特征部分，研究者进行了探索。如文献 [33] 尝试从被动 DNS 记录中提取特征，而文献 [84] 提出了在网络钓鱼电子邮件中应用说服线索。此外，文献 [85] 提出对网页或电子邮件中的 Logo 信息进行特征挖掘。网络钓鱼是一种社会工程攻击，从社会工程进行特征挖掘也有一定潜力，如行为模式、欺骗手法以及受害者的心理反应等。

4.2.3 基线数据集

数据集对于网络钓鱼检测方法的效果确实起到了关键性作用，但当前存在一些问题，包括数据集的不透明性和缺乏全面、实时更新的检测数据集。一些研究者虽然构建了自己的数据集，但通常只用于作者自己的研究模型，这使得其他研究者难以使用这些数据集进行对比实验。另外，一些研究者使用公开数据集，但这些数据集过于陈旧，导致在实际互联网环境中的检测效果较差。

4.2.4 模型抗规避策略

攻击者规避模型检测的方式多种多样，主要可以分为两个方面：首先，攻击者可以采用伪装和规避策略，以阻止用户获取正确的数据信息。其次，攻击者通过对抗性模型和特征解析等方式获取模型使用的特征，并生成相应的规避策略。一些研究者已经开始进行模型抗规避研究，以提高模型对于攻击规避的抵抗力[44]。同时，结合多个不同类型的模型，如基于列表的模型、基于启发式的模型、机器学习模型、深度学习模型和基于图的模型等，可以提高整体检测性能。

5 结束语

网络钓鱼检测在网络安全领域扮演着至关重要的角色。通过对近 4 年的相关文献进行详细分析，对当前网络钓鱼检测场景以及数据集进行分析，并重点对基于列表的检测方法、基于启发式的检测方法、机器学习检测方法、深度学习检测方法以及基于图的检测方法进行分析。总结了检测方法、检测场景、数据集、人为因素等方面存在的一些挑战，并提出一些未来可研究的方向，为网络钓鱼检测提供了一个多角度的总结性工作。

参 考 文 献

[1] Anti-phishing working group. Phishing activity trends report [EB/OL]. [2023-10-20]. https://docs.apwg.org/reports/apwg_trends_report_q4_2022.pdf.

[2] 2023 Data breachinvestigations report[EB/OL]. [2023-10-20]. https://www.verizon.com/business/resources/reports/2023-data-breach-investigations-report-dbir.pdf.

[3] ZHANG P H, OEST A, CHO H, et al. CrawlPhish: Large-scale analysis of client-side cloaking techniques in phishing[C]//Proceedings of the IEEE Symposium on

Security and Privacy. New York: IEEE, 2021: 1109-1124.
[4] SAMARASINGHE N, MANNAN M. On cloaking behaviors of malicious websites[J]. Computers & Security, 2021, 101: 102114.
[5] OEST A, ZHANG P, WARDMAN B, et al. Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale[C]//Proceedings of the 29th USENIX Security Symposium. Boston: USENIX Association, 2020: 361-377.
[6] WEN H X, FANG J Y, WU J J, et al. Hide and seek: An adversarial hiding approach against phishing detection on ethereum[J]. IEEE Transactions on Computational Social Systems, 2023, 10(6): 3512-3523.
[7] ZHANG P H, SUN Z B, KYUNG S, et al. I'm spartacus, no, I'm spartacus: Proactively protecting users from phishing by intentionally triggering cloaking behavior [C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2022: 3165-3179.
[8] OEST A, SAFAEI Y, ZHANG P, et al. PhishTime: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists[C]// Proceedings of the 29th USENIX Security Symposium. Boston: USENIX Association, 2020: 379-396.
[9] ACHARYA B, VADREVU P. PhishPrint: Evading phishing detection crawlers by prior profiling[C]// Proceedings of the 30th USENIX Security Symposium. Vancouver: USENIX Association, 2021: 3775-3792.
[10] LEI W J, HU S Q, HSU C. Unveiling the process of phishing precautions taking: The moderating role of optimism bias[J]. Computers & Security, 2023, 129: 103249.
[11] BAX S, MCGILL T, HOBBS V. Maladaptive behaviour in response to email phishing threats: The roles of rewards and response costs[J]. Computers & Security, 2021, 106: 102278.
[12] SINGH K, AGGARWAL P, RAJIVAN P, et al. Cognitive elements of learning and discriminability in anti-phishing training[J]. Computers & Security, 2023, 127: 103105.
[13] APRUZZESE G, CONTI M, YUAN Y. SpacePhish: The evasion-space of adversarial attacks against phishing website detectors using machine learning[C]//Proceedings of the 38th Annual Computer Security Applications Conference. Austin: ACM, 2022: 171-185.
[14] BITAAB M, CHO H, OEST A, et al. Scam pandemic: How attackers exploit public fear through phishing [C]//Proceedings of the APWG Symposium on Electronic Crime Research. New York: IEEE, 2020: 1-10.
[15] KONDRACKI B, AZAD B A, STAROV O, et al. Catching transparent phish: Analyzing and detecting MITM phishing toolkits[C]//Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2021: 36-50.
[16] CHIEW K L, YONG K S C, TAN C L. A survey of phishing attacks: Their types, vectors and technical approaches[J]. Expert Systems with Applications, 2018, 106: 1-20.
[17] BHATTACHARYA M, ROY S, CHATTOPADHYAY S, et al. ASPA-MOSN: An efficient user authentication scheme for phishing attack detection in mobile online social networks[J]. IEEE Systems Journal, 2023, 17(1): 234-245.
[18] SHAH A, CHANDRAN N, DEMA M, et al. Secure featurization and applications to secure phishing detection [C]//Proceedings of the 2021 on Cloud Computing Security Workshop. New York: ACM, 2021: 83-95.
[19] BARRACLOUGH P A, FEHRINGER G, WOODWARD J. Intelligent cyber-phishing detection for online[J]. Computers & Security, 2021, 104: 102123.
[20] BACANIN N, ZIVKOVIC M, ANTONIJEVIC M, et al. Addressing feature selection and extreme learning machine tuning by diversity-oriented social network search: An application for phishing websites detection[J]. Complex & Intelligent Systems, 2023, 9(6): 7269-7304.
[21] SAKA T, VANIEA K, KÖKCIYAN N. Context-based clustering to mitigate phishing attacks[C]//Proceedings of the 15th ACM Workshop on Artificial Intelligence and Security. New York: ACM, 2022: 115-126.
[22] AZEEZ N A, MISRA S, MARGARET I A, et al. Adopting automated whitelist approach for detecting phishing attacks[J]. Computers & Security, 2021, 108: 102328.
[23] RAO R S, PAIS A R, ANAND P. A heuristic technique to detect phishing websites using TWSVM classifier[J]. Neural Computing and Applications, 2021, 33(11): 5733-5752.
[24] DA SILVA C M R, FERNANDES B J T, FEITOSA E L, et al. Piracema. io: A rules-based tree model for phishing prediction[J]. Expert Systems with Applications, 2022, 191: 116239.
[25] TAN C C L, CHIEW K L, YONG K S C, et al. Hybrid phishing detection using joint visual and textual identity[J]. Expert Systems with Applications, 2023, 220: 119723.
[26] LIU D J, GENG G G, JIN X B, et al. An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment[J]. Computers & Security, 2021, 110: 102421.
[27] TRINH N B, PHAN T D, PHAM V H. Leveraging deep learning image classifiers for visual similarity-based phishing website detection[C]//Proceedings of the 11th International Symposium on Information and Communication Technology. New York: ACM, 2022: 134-141.
[28] CHAI Y D, ZHOU Y H, LI W F, et al. An explainable multi-modal hierarchical attention model for developing phishing threat intelligence[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(2): 790-803.
[29] LIANG Y J, WANG Q S, XIONG K, et al. Robust detection of malicious URLs with self-paced wide & deep learning[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(2): 717-730.
[30] TAN C L, CHIEW K L, YONG K S C, et al. A graph-theoretic approach for the detection of phishing webpages[J]. Computers & Security, 2020, 95: 101793.

[31] FU B X, YU X, FENG T. CT-GCN: A phishing identification model for blockchain cryptocurrency transactions[J]. International Journal of Information Security, 2022, 21(6): 1223-1232.
[32] ZENG Y W, LIU Z C, CHEN X X, et al. Hidden path: Understanding the intermediary in malicious redirections[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1725-1740.
[33] ALSABAH M, NABEEL M, BOSHMAF Y, et al. Content-agnostic detection of phishing domains using certificate transparency and passive DNS[C]//Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses. New York: ACM, 2022: 446-459.
[34] 樊昭杉, 王青, 刘俊荣, 等. 域名滥用行为检测技术综述[J]. 计算机研究与发展, 2022, 59(11): 2581-2605.
FAN Z S, WANG Q, LIU J R, et al. Survey on domain name abuse detection technology[J]. Journal of Computer Research and Development, 2022, 59(11): 2581-2605.
[35] BOZKIR A S, DALGIC F C, AYDOS M. Grambeddings: A new neural network for URL based identification of phishing web pages through N-gram embeddings[J]. Computers & Security, 2023, 124: 102964.
[36] PRABAKARAN M K, MEENAKSHI S P, CHANDRASEKAR A D. An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders[J]. IET Information Security, 2023, 17(3): 423-440.
[37] XIAO X, XIAO W T, ZHANG D Y, et al. Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets[J]. Computers & Security, 2021, 108: 102372.
[38] RAO R S, VAISHNAVI T, PAIS A R. CatchPhish: Detection of phishing websites by inspecting URLs[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(2): 813-825.
[39] DESOLDA G, FERRO L S, MARRELLA A, et al. Human factors in phishing attacks: A systematic literature review[J]. ACM Computing Surveys, 2022, 54(8): 1-35.
[40] LIU R, LIN Y, YANG X, et al. Inferring phishing intention via webpage appearance and dynamics: A deep vision based approach[C]// Proceedings of the 31th USENIX Security Symposium. Vancouver: USENIX Association, 2022: 1633-1650.
[41] GALLO L, MAIELLO A, BOTTA A, et al. 2 Years in the anti-phishing group of a large company[J]. Computers & Security, 2021, 105: 102259.
[42] MCGAHAGAN J, BHANSALI D, PINTO-COELHO C, et al. Discovering features for detecting malicious websites: An empirical study[J]. Computers & Security, 2021, 109: 102374.
[43] SÁNCHEZ-PANIAGUA M, FIDALGO E, ALEGRE E, et al. Phishing websites detection using a novel multipurpose dataset and web technologies features[J]. Expert Systems with Applications, 2022, 207: 118010.
[44] KIM T, PARK N, HONG J, et al. Phishing URL detection: A network-based approach robust to evasion[C]// Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2022: 1769-1782.
[45] YOO J, CHO Y. ICSA: Intelligent chatbot security assistant using Text-CNN and multi-phase real-time defense against SNS phishing attacks[J]. Expert Systems with Applications, 2022, 207: 117893.
[46] YANG L Q, ZHANG J W, WANG X Z, et al. An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features[J]. Expert Systems with Applications, 2021, 165: 113863.
[47] BUSTIO-MARTÍNEZ L, ÁLVAREZ-CARMONA M A, HERRERA-SEMENETS V, et al. A lightweight data representation for phishing URLs detection in IoT environments[J]. Information Sciences, 2022, 603: 42-59.
[48] BOUNTAKAS P, XENAKIS C. HELPHED: Hybrid ensemble learning phishing email detection[J]. Journal of Network and Computer Applications, 2023, 210: 103545.
[49] JHA A K, MUTHALAGU R, PAWAR P M. Intelligent phishing website detection using machine learning[J]. Multimedia Tools and Applications, 2023, 82(19): 29431-29456.
[50] ALANI M M, TAWFIK H. PhishNot: A cloud-based machine-learning approach to phishing URL detection[J]. Computer Networks, 2022, 218: 109407.
[51] ZHU E Z, CHEN Z L, CUI J, et al. MOE/RF: A novel phishing detection model based on revised multiobjective evolution optimization algorithm and random forest[J]. IEEE Transactions on Network and Service Management, 2022, 19(4): 4461-4478.
[52] 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]. Information Sciences, 2019, 484: 153-166.
[53] NAGUNWA T, KEARNEY P, FOUAD S. A machine learning approach for detecting fast flux phishing hostnames[J]. Journal of Information Security and Applications, 2022, 65: 103125.
[54] GUPTA B B, YADAV K, RAZZAK I, et al. A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment[J]. Computer Communications, 2021, 175: 47-57.
[55] LEE J, YE P X, LIU R F, et al. Building robust phishing detection system: An empirical analysis[C]//Proceedings of the 2020 Workshop on Measurements, Attacks, and Defenses for the Web. San Diego: ISOC, 2020: 1-12.
[56] DA S C M R, FEITOSA E L, GARCIA V C. Heuristic-based strategy for phishing prediction: A survey of URL-based approach[J]. Computers & Security, 2020, 88: 101613.
[57] VAITKEVICIUS P, MARCINKEVICIUS V. Comparison of classification algorithms for detection of phishing websites[J]. Informatica, 2020: 143-160.
[58] SERN L J, PENG DAVID Y G, HAO C J. PhishGAN: Data augmentation and identification of homoglyph attacks[C]//Proceedings of the International Conference on Communications, Computing, Cybersecurity, and Informatics. New York: IEEE, 2020: 1-6.

[59] VALENTIM R, DRAGO I, TREVISAN M, et al. Augmenting phishing squatting detection with GANs[C]// Proceedings of the CoNEXT Student Workshop. New York: ACM, 2021: 3-4.
[60] WEI W, KE Q, NOWAK J, et al. Accurate and fast URL phishing detector: A convolutional neural network approach[J]. Computer Networks, 2020, 178: 107275.
[61] HUSSAIN M, CHENG C, XU R, et al. CNN-Fusion: An effective and lightweight phishing detection method based on multi-variant ConvNet[J]. Information Sciences, 2023, 631: 328-345.
[62] ZHENG F A, YAN Q, LEUNG V C M, et al. HDP-CNN: Highway deep pyramid convolution neural network combining word-level and character-level representations for phishing website detection[J]. Computers & Security, 2022, 114: 102584.
[63] DHANAVANTHINI P, CHAKKRAVARTHY S S. Phish-armour: Phishing detection using deep recurrent neural networks[J]. Soft Computing, 2023, 11: 7962.
[64] XIAO X, ZHANG D Y, HU G W, et al. CNN–MHSA: A convolutional neural network and multi-head self-attention combined approach for detecting phishing websites[J]. Neural Networks, 2020, 125: 303-312.
[65] WANG C G, CHEN Y Y. TCURL: Exploring hybrid transformer and convolutional neural network on phishing URL detection[J]. Knowledge-Based Systems, 2022, 258: 109955.
[66] WAZIRALI R, AHMAD R, ABU-EIN A A K. Sustaining accurate detection of phishing URLs using SDN and feature selection approaches[J]. Computer Networks, 2021, 201: 108591.
[67] LIN Y, LIU R F, DIVAKARAN D, et al. Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages[C]//Proceedings of the 30th USENIX Security Symposium. Vancouver: USENIX Association, 2021: 3793-3810.
[68] ABDELNABI S, KROMBHOLZ K, FRITZ M. VisualPhishNet: Zero-day phishing website detection by visual similarity[C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2020: 1681-1698.
[69] LI Q, CHENG M Y, WANG J F, et al. LSTM based phishing detection for big email data[J]. IEEE Transactions on Big Data, 2022, 8(1): 278-288.
[70] WEN T K, XIAO Y X, WANG A Q, et al. A novel hybrid feature fusion model for detecting phishing scam on Ethereum using deep neural network[J]. Expert Systems with Applications, 2023, 211: 118463.
[71] NABEEL M, KHALIL I M, GUAN B, et al. Following passive DNS traces to detect stealthy malicious domains via graph inference[J]. ACM Transactions on Privacy and Security, 2020, 23(4): 1-36.
[72] ALHOGAIL A, ALSABIH A. Applying machine learning and natural language processing to detect phishing email[J]. Computers & Security, 2021, 110: 102414.
[73] CHEN L, PENG J Y, LIU Y, et al. Phishing scams detection in ethereum transaction network[J]. ACM Transactions on Internet Technology, 2021, 21(1): 1-16.
[74] LIN X, ILIA P, SOLANKI S, et al. Phish in sheep's clothing: Exploring the authentication pitfalls of browser fingerprinting[C]//Proceedings of the 31th USENIX Security Symposium . Vancouver: USENIX Association, 2022: 1651-1668.
[75] ULQINAKU E, ASSAL H, ABDOU A R, et al. Is real-time phishing eliminated with FIDO? Social engineering downgrade attacks against FIDO protocols[C]// Proceedings of the 30th USENIX Security Symposium. Vancouver: USENIX Association, 2021: 3811-3828.
[76] HU H, JAN S T K, WANG Y, et al. Assessing browser-level defense against IDN-based phishing[C]// Proceedings of the 30th USENIX Security Symposium. Vancouver: USENIX Association, 2021: 3739-3756.
[77] CHEN Y, ZAHEDI F M, ABBASI A, et al. Trust calibration of automated security IT artifacts: A multi-domain study of phishing-website detection tools[J]. Information & Management, 2021, 58(1): 103394.
[78] HULL D M, SCHUETZ S W, LOWRY P B. Tell me a story: The effects that narratives exert on meaningful-engagement outcomes in antiphishing training[J]. Computers & Security, 2023, 129: 103252.
[79] NGUYEN C, JENSEN M, DAY E. Learning not to take the bait: A longitudinal examination of digital training methods and overlearning on phishing susceptibility[J]. European Journal of Information Systems, 2023, 32(2): 238-262.
[80] DRURY V, ROEPKE R, SCHROEDER U, et al. Analyzing and creating malicious URLs: A comparative study on anti-phishing learning games[C]//Proceedings of the 2022 Symposium on Usable Security. Reston, VA: Internet Society, 2022: 14-27.
[81] JAYAKRISHNAN G, BANAHATTI V, LODHA S. PickMail: A serious game for email phishing awareness training[C]//Proceedings of the 2022 Symposium on Usable Security. Reston, VA: Internet Society, 2022: 1-13.
[82] REINHEIMER B, ALDAG L, MAYER P, et al. An investigation of phishing awareness and education over time: When and how to best remind users[C]//Proceedings of the Sixteenth Symposium on Usable Privacy and Security. San Francisco: USENIX Association, 2020: 259-284.
[83] ALTHOBAITI K, MENG N, VANIEA K. I don't need an expert! Making URL phishing features human comprehensible[C]//Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2021: 1-17.
[84] VALECHA R, MANDAOKAR P, RAO H R. Phishing email detection using persuasion cues[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(2): 747-756.
[85] BOZKIR A S, AYDOS M. LogoSENSE: A companion HOG based logo detection scheme for phishing web page and E-mail brand recognition[J]. Computers & Security, 2020, 95: 101855.

编　辑　张　莉