TxPhishScope: Towards Detecting and Understanding Transaction-based Phishing on Ethereum

Bowen He*
Zhejiang University
China
bowen_os@zju.edu.cn

Yuan Chen*
Zhejiang University
China
yuanchen96@zju.edu.cn

Zhuo Chen*
Zhejiang University
China
hypothesiser.hypo@zju.edu.cn

Xiaohui Hu*
Beijing University of Posts and Telecommunications
China
hxhpx@bupt.edu.cn

Yufeng Hu*
Zhejiang University
China
yufenghu@zju.edu.cn

Lei Wu†
Zhejiang University
China
lei_wu@zju.edu.cn

Rui Chang
Zhejiang University
China
crix1021@zju.edu.cn

Haoyu Wang
Huazhong University of Science and Technology
China
haoyuwang@hust.edu.cn

Yajin Zhou†‡
Zhejiang University
China
yajin_zhou@zju.edu.cn

ABSTRACT

The prosperity of Ethereum attracts many users to send transactions and trade crypto assets. However, this has also given rise to a new form of transaction-based phishing scam, named TxPhish. Specifically, tempted by high profits, users are tricked into visiting fake websites and signing transactions that enable scammers to steal their crypto assets. The past year has witnessed 11 large-scale TxPhish incidents causing a total loss of more than $70 million.

In this paper, we conduct the first empirical study of TxPhish on Ethereum, encompassing the process of a TxPhish campaign and details of phishing transactions. To detect TxPhish websites and extract phishing accounts automatically, we present TxPhishScope, which dynamically visits the suspicious websites, triggers transactions, and simulates results. Between November 25, 2022, and July 31, 2023, we successfully detected and reported 26,333 TxPhish websites and 3,486 phishing accounts. Among all of documented TxPhish websites, 78.9% of them were first reported by us, making TxPhishScope the largest TxPhish website detection system. Moreover, we provided criminal evidence of four phishing accounts and their fund flow totaling $1.5 million to aid in the recovery of funds for the victims. In addition, we identified bugs in six Ethereum projects and received appreciation.

Based on the detection results, we perform a comprehensive study of TxPhish websites and phishing accounts. Our study reveals that TxPhish websites have a short lifespan, low cost, and fast update frequency. Besides, Our analysis of phishing fund flow demonstrates that 54.0% of phishing funds ($43.7 million) flowed into centralized exchanges, where the identity of owners could be traced. Our research can serve as a valuable reference for Ethereum service providers to safeguard their users against TxPhish and assist in the recovery of victims' crypto assets.

CCS CONCEPTS

• **Security and privacy → Web application security**.

KEYWORDS

Decentralized Finance, Ethereum, phishing detection

ACM Reference Format:
Bowen He, Yuan Chen, Zhuo Chen, Xiaohui Hu, Yufeng Hu, Lei Wu, Rui Chang, Haoyu Wang, and Yajin Zhou. 2023. TxPhishScope: Towards Detecting and Understanding Transaction-based Phishing on Ethereum. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23), November 26–30, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3576915.3623210

1 INTRODUCTION

Recent years have seen a significant growth of Decentralized Finance (DeFi) based on blockchain. Ethereum is the second generation of blockchain platforms, which supports both peer-to-peer transactions and program execution, in the form of smart contracts [67]. Based on these properties, developers can issue tokens [1] and build decentralized applications (DApps). Due to these features,

*Part of the work was finished when the authors were research interns at BlockSec.
†These authors are also affiliated at Key Laboratory of Blockchain and Cyberspace Governance of Zhejiang Province.
‡Yajin Zhou is the corresponding author.

CCS '23, November 26–30, 2023, Copenhagen, Denmark
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0050-7/23/11…$15.00
https://doi.org/10.1145/3576915.3623210

[1] We use the term "token" to denote users' crypto assets on a blockchain.

Ethereum is attracting significant investments from a large number of users. As of August 2023, the Total Value Locked (TVL) in Ethereum has exceeded $24.1 billion, representing 58.1% of the total TVL [60] in the DeFi sector.

Definition of TxPhish. As users trade tokens by sending transactions on Ethereum, a new form of phishing scam has emerged. Unlike traditional ones that target victims' personal or financial information [74, 84, 85, 96, 97], this type of phishing steals users' assets through transactions. Specifically, scammers trick victims into signing transactions or messages [2] that enable them to withdraw the victims' tokens in transactions. Since the phishing scam mainly involves victims' signing transactions, we name it TxPhish in this paper. Between February 2022 and April 2023, we have witnessed 11 significant TxPhish incidents [8–10, 14, 15, 31–33, 38, 56, 64], which led to a collective loss of more than $70 million.

The whole process of a TxPhish campaign generally consists of several steps. To begin with, scammers lure users to visit phishing websites and connect their wallets. The scam program then searches for valuable tokens in the user's account and creates a phishing transaction. Users are led to believe that the transaction can generate profits and blindly sign it. Consequently, their tokens are swiftly drained.

Existing solutions to combat TxPhish. To fight phishing scams on Ethereum, several Web3 service providers [37, 44] and security organizations [25, 39] have set up lists of phishing domains and accounts. The source of these lists mainly relies on users' reports, which can be divided into three categories. The first type is victims' evidence of being deceived [54]. The second type consists of fake advertising on social platforms [52]. The third type is from Web3 security researchers manually searching for websites that share the same code hash features with known phishing sites [53].

However, the above methods cannot block a significant number of TxPhish websites in time. First, since phishing websites would go offline quickly after being spread out, the former two methods are too late to block them. Consequently, many websites have been taken down by the time they are reported. Second, as the toolkit of TxPhish websites is continually and rapidly updated, the code hash feature mapping can be bypassed easily, which makes the third method fall short in the detection scale of TxPhish websites. Our collected data shows that, from September 1, 2022, to November 1, 2022, only around 34 TxPhish websites were reported per day in Chainabuse [26], the state-of-the-art Web3 anti-phishing platform. Third, *existing methods are all based on manual work*. Since scammers can launch phishing websites at any time, it is important to deploy a detection system that can identify them promptly. Besides, *the information of phishing accounts is neglected in most cases*. The above three methods all focus on detecting and reporting domains of TxPhish websites. Only those phishing accounts deceiving millions of dollars would be reported in some cases. Nevertheless, uncovering phishing accounts is helpful for safeguarding users and recovering funds for victims. In summary, *an automated system that can detect a large number of TxPhish websites timely and extract phishing accounts simultaneously* is in need.

TxPhishScope. Although the phishing websites are evolving constantly, we observe that the process of TxPhish is generally fixed. So we propose a system to detect TxPhish websites by *dynamically triggering and simulating transactions*, named as TxPhishScope. To begin with, it retrieves domains from real-time registered certificates via Certificate Transparency (CT) [22]. Then, it assesses whether these domains are related to Ethereum by analyzing both the domain and web page content. Next, TxPhishScope dynamically visits the target website, connects to the wallet, and clicks on the token trade button to trigger a transaction. Finally, it simulates the execution of the transaction [47] and employs rigorous logic rules to determine whether it is a phishing transaction. When detecting phishing transactions, TxPhishScope would automatically save the evidence chain, including screenshots, source codes of websites, and details of phishing transactions for reporting purposes.

We have deployed TxPhishScope to perform a large-scale detection of TxPhish websites for over eight months. To block detected TxPhish websites and prevent users' losses, we verify the evidence chain and report them to two reputable entities: MetaMask [44], the most popular Ethereum wallet, and Forta [39], a real-time detection network for security monitoring of blockchain activity. Additionally, we report the phishing accounts to Etherscan [37], the most popular browser on Ethereum. From November 25, 2022, to July 31, 2023, we reported **26,333** TxPhish websites and **3,486** phishing accounts with no false positives. Among them, **93.1%** of TxPhish websites are first discovered by us. Our collected data from MetaMask and Chainabuse shows that **78.9%** of all TxPhish websites are first reported by us. To the best of our knowledge, TxPhishScope is the largest real-time detection system of TxPhish websites. What's more, we provided criminal evidence of **four** phishing accounts and their fund flow of **$1.5 million** to aid in the recovery of funds for victims. Besides, our system has also helped to identify bugs in **six** Ethereum projects, avoiding potential user losses and receiving appreciation.

Measurements. After understanding the implementation details of TxPhishScope, scammers may try to develop specific techniques aimed at bypassing our detection mechanisms. Nevertheless, we have gathered a significant volume of data for the purpose of measurements. To gain further insights into the ecosystem, we conduct a comprehensive analysis of TxPhish websites from three different aspects, namely lifespan, cost, and update frequency. Our observations reveal that the average online duration of a TxPhish website is **113** hours. Although the website would go offline quickly, the same phishing account may appear in another one. Statistically, we identify **445** phishing accounts that appear across different websites for over a month. To save costs, scammers prefer to register free certificates and host multiple websites under the same parent domain. We have discovered **6,754** such websites that share **673** parent domains. Shockingly, the cost of setting up a phishing website can be as low as **$0.13**. Furthermore, given the rapid emergence of new Web3 projects and technologies, the imitation targets of TxPhish websites change accordingly with the hot events (from TrustPad [61], Blur [21], zksync [70], Arbitrum [16], AIDOGE [12], to pepe [50]).

Finally, to assist in the recovery of funds for victims, we conduct a comprehensive analysis of phishing fund flow. To achieve this, we summarize common techniques used to handle phishing funds,

[2] For simplicity, we would use the term "phishing transactions" to refer to both phishing messages and transactions, which users are deceived to sign, in the following sections.

including exchanging through centralized exchanges (CEXs) and laundering money via mixers or cross-chain bridges. In addition, we collect **10,705** incoming transactions for **630** active phishing accounts and build a fund flow graph for each account. Unlike Ethereum attacks causing millions of dollars to be lost, phishing scams typically only earn a few thousand dollars from a single account. Although the identity of the fund owners in CEXs could be traced, it is challenging for all victims to collect their evidence and jointly contact the authorities. So scammers prefer to directly exchange phishing funds via CEXs. Our results show that of the **$80.8 million** phishing funds, **54.0%** flowed into centralized exchanges, while only **14.7%** flowed into mixers or cross-chain bridges. We also discover several accounts that received large amounts of phishing funds, which remain unprocessed as of August 10, 2023. Furthermore, we observe that scammers transferred **$2.7 million** to an online casino service [58], which is also a form of money laundering that is difficult to trace.

Contributions. Despite extensive research efforts to detect and measure phishing websites [72, 74–76, 84, 88–90, 92–97, 100, 102, 107, 109], to the best of our knowledge, we have not found any existing literature that specifically addresses TxPhish websites on Ethereum. Our work aims to fill this gap. In addition, TxPhishScope can be deployed continuously to detect more TxPhish websites and reveal more attributes of them. We hope that our work can serve as a guide to help Ethereum service providers protect their users from phishing scams and aid in the recovery of victims' tokens. Our contributions are summarized as follows.

- We first systematize TxPhish on Ethereum, covering both the process of a TxPhish campaign and details of phishing transactions.
- We build the largest TxPhish website detection system, named TxPhishScope. Our collected data shows that **78.9%** of TxPhish websites are first reported by us.
- We conduct a comprehensive measurement of TxPhish websites, focusing on their lifespan, cost, and update frequency.
- We summarize common techniques used for handling phishing funds and make a thorough analysis of fund flow targets.

2 BACKGROUND

2.1 Ethereum Basics

Ethereum is a decentralized blockchain platform [67] that enables developers to deploy decentralized applications (DApps). It uses Ether (ETH) as its native token to facilitate transactions and pay for computations. Ethereum is currently the second-largest blockchain network by market capitalization, following Bitcoin.

Accounts. There are two types of accounts [73] in Ethereum: Externally Owned Accounts (EOAs) and Contract Accounts (CAs). EOAs are controlled by private keys, which allow users to access their funds and execute transactions. In contrast, CAs are controlled by the code of their associated smart contracts, which define the rules and logic for executing transactions. Note that, both types of accounts can hold ETH and other tokens.

Transactions. In Ethereum, transactions are signed messages used to transfer tokens, invoke smart contract functions, or deploy smart contracts. Essentially, Ethereum is a global state machine [73] that can be modified by these transactions.

Table 1: Approval and Transfer Functions [3]of ERC-20, ERC-721, and ERC-1151 Interfaces [4]. Approval functions grant permission for the transfer of a user's tokens to another account, while Transfer functions actually move the user's tokens to the designated account.

		Approval Functions	**Transfer Functions**
	ERC-20	*Approve (spender, value)*	*TransferFrom (from, to, value)*
NFT	ERC-721	*Approve (approved, tokenId)*	*TransferFrom (from, to, tokenId)*
	ERC-1155	*SetApprovalForAll (operator, approved)*	*SafeTransferFrom (from, to, id, value, data)*

Smart contracts. Smart contracts in Ethereum are immutable programs that run on the blockchain. They consist of a set of functions and data, which are stored at a unique Ethereum address. To interact with a smart contract, a user can initiate transactions that contain specific parameters to execute its functions.

Wallets. Within the Web3 cryptocurrency ecosystem, users typically rely on digital wallets to manage their private keys and launch transactions. These wallets can be categorized into two types: software wallets and hardware wallets. Software wallets encompass desktop, mobile, and web extension wallets. Among them, MetaMask [44] is the most widely used Ethereum wallet with over 30 million monthly active users [5].

2.2 Ethereum Tokens

Ethereum allows developers to create and issue tokens, which can be used for purposes like fundraising, rewards, or voting. Specifically, a token is a unit of value created and managed with smart contracts. These tokens are tracked using a mapping in the smart contract code, which records the relationship between accounts and tokens. And tokens can be categorized into fungible (each token is equivalent to the others) and non-fungible tokens (each token is unique and distinct). Generally, the fundamental functions for managing tokens can be divided into two categories: approval and transfer functions. As can be seen from Table 1, approval functions authorize another account to control user tokens, while transfer functions move user tokens to a different account.

ERC-20 Tokens. Most of the fungible token contracts follow the ERC-20 [1] interface, such as USDT, USDC, or WETH. Since each token is identical to the others, the parameter *value* represents how many tokens users can authorize or transfer.

Non-Fungible Tokens (NFT). Currently, most NFT contracts utilize the ERC-721 [3] and ERC-1155 [2] interfaces. ERC-721 Tokens are unique and distinct digital assets that are typically utilized to represent collectibles or artwork. While in the case of ERC-1155 Tokens, NFTs with the same ID are identical and can be used to represent a variety of assets, including in-game items or reward points.

[3] For simplicity, we provide just a few functions as examples.

[4] "Ethereum Requests for Comments (ERCs)" are technical proposals and standards created by developers to define the functionality and interfaces of smart contracts in the Ethereum ecosystem.

2.3 Phishing Scams on Ethereum

With the rapid development of the Ethereum economy, various types of phishing scams have emerged. These scams can appear in various forms, including fake websites, emails, apps [6], and social media messages that mimic legitimate Ethereum services and platforms. Then users are tricked into revealing their mnemonics [7] or signing transactions [4] to transfer their funds. Among these scams, TxPhish happens almost every day [36] and causes millions of dollars in losses [11].

2.4 Anti-phishing Measures on Ethereum

To combat phishing scams on Ethereum, MetaMask has established a list of phishing domains [36]. Users will be restricted from accessing phishing websites recorded in the list, in which case a blocking page would be displayed. Besides, Users can search for and report phishing domains in Chainabuse [26], which is the largest Web3 anti-phishing platform for recording and tracking fraudulent cryptocurrency scams. Meanwhile, Forta has developed a scam detector [40] designed to gather phishing reports and allow users to query labels associated with Ethereum accounts or URLs. In addition, Etherscan [37], the most popular browser on Ethereum, also labels phishing accounts to caution users about the associated risks.

At present, Ethereum phishing reports can be divided into three categories. The first category consists of reports submitted by victims of phishing [54], who provide evidence of having been deceived. The second type contains fake advertising on social platforms such as Twitter, Instagram, and Discord, where both phishing and legitimate project websites are provided for comparison [52]. The third category is derived from Web3 security researchers manually searching for fake official websites that share the same code features with known phishing sites [53]. Upon discovering sites that employ the same phishing toolkits with urlscan [62], the researchers would report them as phishing sites.

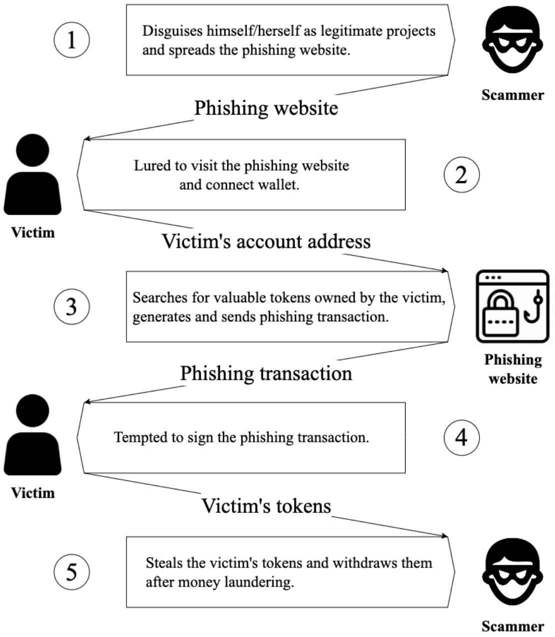
2.5 Limitations of Existing Solutions

However, the aforementioned approaches cannot block a significant number of TxPhish websites in time. Besides, phishing accounts are also overlooked in most cases.

Many websites have gone offline when reported. The first two methods are too late to prevent the proliferation of TxPhish websites, as they have been spread out widely. Moreover, sometimes the websites have already been taken down when they are reported. For these cases, scammers may have launched new TxPhish websites and would likely disseminate them at once.

The number of reported TxPhish websites is limited. Although the last approach identifies TxPhish websites based on their static code features, it can still be easily bypassed by applying a different phishing toolkit. To illustrate this point, we have collected statistics on TxPhish websites reported in Chainabuse [26] from September 1, 2022, to November 1, 2022. We find that there were only 2,081 TxPhish website reports, around 34 per day, which is significantly lower than the number of websites detected by TxPhishScope.

Detecting TxPhish websites are all based on manual work. All of the above-mentioned methods rely on individuals' efforts to manually identify TxPhish websites. However, scammers can launch them at any moment, which highlights the need for an automated system that can detect TxPhish websites at any time.



1
Disguises himself/herself as legitimate projects and spreads the phishing website.
Scammer
Phishing website
Lured to visit the phishing website and connect wallet.
2
Victim
Victim's account address
3
Searches for valuable tokens owned by the victim, generates and sends phishing transaction.
Phishing website
Phishing transaction
Tempted to sign the phishing transaction.
4
Victim
Victim's tokens
5
Steals the victim's tokens and withdraws them after money laundering.
Scammer

Figure 1: Workflow of a TxPhish Campaign

The information of phishing accounts is neglected. Currently, the majority of phishing reports center solely on domains of TxPhish websites, neglecting to uncover the actual phishing transactions. Yet, it is the phishing account that ultimately receives all the tokens from victims. It's common to find the same phishing account repeatedly surfacing on various websites. Consequently, keeping a record of these phishing accounts can prove to be an effective approach in safeguarding users from TxPhish websites. Furthermore, by examining a phishing account's history transactions, we can gain insights into the flow of funds and potentially assist victims in recovering their lost funds.

3 ANATOMY OF TXPHISH

In order to gain a preliminary understanding of the TxPhish campaigns, we select 11 large-scale TxPhish events mentioned above [8–10, 14, 15, 31–33, 38, 56, 64], which resulted in a total loss of more than $70 million. Furthermore, we have gathered a total of 642 TxPhish incident reports from several Web3 security companies [19, 20, 23]. In this section, we would first describe the entire process of a TxPhish campaign. Then, we would analyze the types of phishing transactions and the signing methods used when victims visit these websites. These efforts can serve as a theoretical basis for detecting TxPhish websites.

3.1 Process of a TxPhish Campaign

As depicted in Figure 1, the entire process of a TxPhish campaign consists of five steps. We would provide a thorough description of them in the following paragraphs.

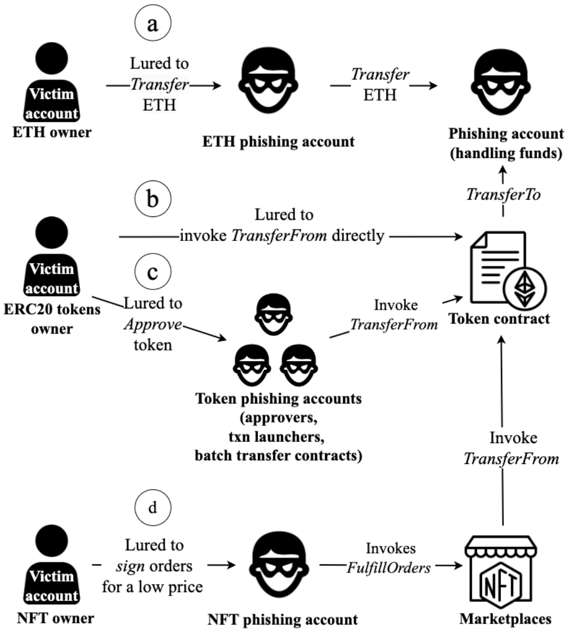


Figure 2: Types of phishing transactions

Scammer spreads fake websites. Generally, the scammer disseminates fraudulent websites on social platforms like Twitter, Telegram, Instagram, or Discord. Additionally, they may inject malicious advertisements containing phishing websites into Google Search [31]. To reach more users, the scammer would impersonate well-known projects or compromise their official accounts directly. And we discover that 24.6% of phishing incidents resulted from popular project accounts being hacked.

Victim clicks and connects wallets. Attracted by words like "FREE CLAIM", the victim would visit the phishing website without in-depth thinking. Since many fraudulent websites are generated with website copy tools, they appear identical to legitimate ones. As a result, the victim connects the wallet quickly.

Phishing website creates a phishing transaction or message. Once the victim's account address is obtained, the frontend program in the phishing website uses free APIs, like OpenSea [49], Moralis [48], and Alchemy [13], to scan for valuable tokens. Following this, a phishing transaction or message is created to steal the victim's most prized tokens.

Victim signs the phishing transaction or message. As the phishing website is well disguised, the victim fails to distinguish it from the legitimate one and mistakenly signs the phishing transaction. Sometimes, the scammer entices the victim to sign a message that enables them to withdraw tokens via transactions.

Scammer withdraws the victim's tokens and cashes them out. After receiving the victim's tokens, the scammer would cash them out from centralized cryptocurrency exchanges, like Binance [18] or Coinbase [27]. In some cases, to conceal the ultimate destination of the victim's tokens, the scammer would first launder these tokens.

3.2 Analysis of Phishing Transactions

As previously stated, after retrieving the victim's account address, the phishing website proceeds to send a phishing transaction. In addition, the parameters of the phishing transaction are determined by the value of the victim's tokens. Specifically, if it is unfeasible to withdraw all tokens belonging to users, the phishing transaction would prioritize those with the highest value. And a detailed analysis of phishing transactions is as follows.

3.2.1 Types of Phishing Transactions. Figure 2 provides a summary of phishing transactions for different victims, including transfer, approval, and zero-dollar purchase phishing. Generally, victims' tokens would be transferred to a phishing account for funds handling in Section 6.

Transfer phishing. In transfer phishing, the victim is tempted to transfer ETH or ERC20 tokens to a phishing account, which corresponds to type **a** and **b** illustrated in Figure 2. The former is a transaction directly sending ETH to the phishing account. While the latter one is a transaction sending ERC20 tokens to the phishing account via invoking *TransferFrom* in the token contract.

Approval phishing. Type **c** in Figure 2 depicts the process of approval phishing. Specifically, The victim is lured to authorize a phishing account to control his or her tokens via approval functions like *Approve* or *Permit*. Then scammers invoke *TransferFrom* to move the victim's funds. In addition, to save gas fees, some scammers deploy a contract to transfer multiple victims' tokens in one transaction. So approval phishing often involves two or more phishing accounts.

Zero-dollar purchase phishing. Typically, users buy and sell NFTs in marketplaces such as OpenSea [49] or Blur [21]. They trade NFTs by signing orders and launching transactions. In zero-dollar purchase phishing, the victim is deceived into signing an order for a low price or even for free. With the order signature, the scammer would launch transactions and fulfill orders to transfer the victim's tokens quickly. The phishing process is described in Type **d**.

3.2.2 Signing Methods of Phishing Transactions. MetaMask [44] offers several APIs for requesting signatures from users. Some of them are exploited in phishing transactions to trick unsuspecting victims.

eth_sendTransactions. This is the most basic way to sign and send transactions on Ethereum. Users can view all the transaction parameters in the popup window, including the target address of the transaction, the amount of ETH sent, and function parameters if it interacts with a contract. In contrast, the following two methods are invoked to sign a piece of data. Then scammers need to manually launch transactions with the victims' signatures to withdraw their tokens.

eth_signTypedData. This method allows requesting users to sign a readable data structure that can be verified on-chain [46]. Users can see details of the struct in the popup window. And it is typically invoked when users need to sign an order in marketplaces like OpenSea [49] or Blur [21].

eth_sign. This method asks users to sign an arbitrary message, which is a hexadecimal sequence in the popup window. Nonetheless, as signing a transaction is equivalent to signing its hash, a

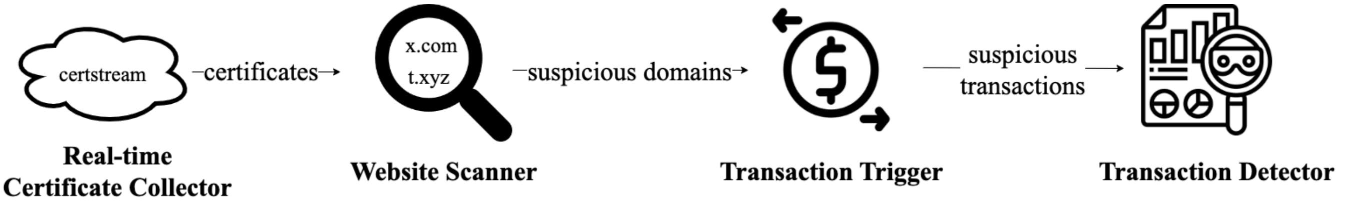


Figure 3: Components of TxPhishScope

phishing website can directly prompt users to sign the hash of a phishing transaction. Since users are unable to recover transaction parameters from the hash, they are unaware of the transaction's intended behavior. Due to its exploitation by scammers on phishing websites, MetaMask has disabled eth_sign by default [45].

4 TXPHISHSCOPE

As mentioned in Section 2.5, there lacks an automated system that can promptly detect large numbers of TxPhish websites and extract phishing transactions. In order to detect and report TxPhish websites effectively, we build a detection system, named TxPhishScope. Figure 3 illustrates the overall architecture of TxPhishScope, including a real-time certificate collector, website scanner, transaction trigger, and transaction detector.

4.1 Real-time Certificate Collector

In order to prevent the broadcast of phishing websites on social media, it is crucial to detect and block them as soon as possible. Based on reports of the Anti-Phishing Working Group [51], Over 80% phishing websites use the HTTPS protocol. Generally, scammers would deploy and launch phishing websites rapidly [74]. So we propose to retrieve suspicious domains from real-time issued certificates. To this end, we have deployed a certstream server [24] to receive real-time updates from the Certificate Transparency [5] Log network [22].

4.2 Website Scanner

To efficiently sift through vast numbers of certificates and identify potentially fraudulent domains, we first analyze 2,081 phishing website reports collected in Section 2.4. In statistics, 72.8% of phishing website certificates are issued by *Let's Encrypt*. Besides, 40.0% of phishing domains contain appealing Web3-related keywords like "mint", "airdrop", and "nft". Furthermore, a staggering 93.8% of phishing domains mimic the brand name of existing Ethereum projects to lure victims. For comparative analysis, we gather the official websites of the top 2,000 blockchain Dapps from DappRadar [34], which offers comprehensive information about the most popular existing Dapps. Among these legitimate websites, 47.1% of their certificates are issued by *Let's Encrypt*. And only 7.7% of them include the aforementioned appealing keywords. Thus, we develop a feature-based scoring algorithm as our initial strategy for identifying suspicious domains.

Algorithm 1 Domain Scoring Algorithm

```
Input: domain: domain extracted from certificate;
  certificate_issuer: certificate issuer extracted from certificate;
  keywords: Web3-related keywords;
  keyword_score: score related with each keyword;
  project_names: names of collected Ethereum projects;
Output: domain_score
  initialize domain_score = 0;
  if certificate_issuer == "Let's Encrypt" then
      domain_score += 20
  end if
  if domain utilizes Punycode then
      domain_score += 20
      convert domain to its regular form
  end if
  for keyword in keywords do
      if domain contains keyword then
          domain_score += keyword_score[keyword]
      end if
  end for
  initialize max_ratio = 0
  for subdomain in domain do
      for name in projects_names do
          if max_ratio < Levenshtein(subdomain, name) then
              max_ratio = Levenshtein(subdomain, name)
          end if
      end for
  end for
  domain_score += max_ratio*100
  return domain_score
```

Domain Scoring Algorithm. Algorithm 1 demonstrates the domain scoring process. To begin with, we assign the initial score based on whether the website certificate is issued by *Let's Encrypt*. Additionally, we need to identify phishing domains that use Punycode to mimic top projects [74], such as xn–pple-43d.com for apple.com, which are not present in any of the legitimate datasets mentioned above. For these cases, we would convert them to regular domains and add scores correspondingly. Subsequently, we add the score for the presence of Web3-related keywords such as "mint", "airdrop", "nft", and "eth" in the domain. Finally, we determine the similarity ratio between each subdomain and names of common Ethereum projects using the Levenshtein.ratio function, and add the highest ratio multiplying 100 to the score. To this end, we keep crawling names of tokens and projects from etherscan [37], opensea [49], and real-time transactions. Now we have collected 41 keywords and 243,722 project names. Once a threshold of 90 is

[5] Certificate Transparency (CT) is an Internet security standard, through which we can monitor the issue of certificates.

reached, the domain is flagged as suspicious. Our experimental data shows that the domain scoring algorithm can successfully detect 94.8% of the phishing domains collected in Section 2.4.

Website Crawler. However, only a few numbers of the websites flagged above are related to Ethereum, bringing unnecessary time overhead for the following detection. Statistically, in around 35 million certificates received from the certstream server every day, we could detect around two million suspicious domains. To further reduce the number of false positives, we build a crawler to judge whether a target website is related to Ethereum. Since the state of the Ethereum network is constantly changing, any service operating on it must retrieve its current state. As a result, they are required to utilize certain public APIs to query crucial information, such as account state or token value. So our crawler would search for specific API-related keywords, like "moralis", "infura", and "alchemy", in the source code of the target website. In some cases, phishing websites may try to conceal their malicious behavior by saving their phishing code in sublinks. To address this, our crawler is designed to match recursively and thoroughly search all possible sublinks. Assuming the phishing website would be launched in 24 hours, our crawler would visit the website every eight hours during this time period. After this step, the number of suspicious websites can sharply decrease to around 4,000.

4.3 Transaction Trigger

Since the key feature of TxPhish is the phishing transaction, its result can be easily determined by retrieving its parameters. Therefore, we propose to detect TxPhish websites based on their transactions. And we would demonstrate the design of the transaction trigger in this section.

As discussed in Section 3.1, the crucial steps in the TxPhish process involve connecting the wallet and signing the transaction. To swindle victims out of their tokens as soon as possible, the phishing process must be streamlined and rapid. In the majority of cases, the victim initially clicks on a "connect" button to link their wallet. Subsequently, the phishing transaction appears after the victim clicks on the "mint" button. Although the toolkits used by phishing websites frequently change, the workflow of a TxPhish campaign remains fixed.

Drawing from the aforementioned observations, we have developed the transaction trigger utilizing the *puppeteer* library [55]. Specifically, the transaction trigger camouflages itself as an ordinary Web3 user and visits the suspicious website. In the initial phase, it clicks on buttons containing keywords such as "connect" to establish a connection with MetaMask. Subsequently, in the second phase, it clicks on buttons containing keywords such as "mint" or "claim" to execute the phishing transaction. To obtain the transaction parameters, we intercept all signing methods provided by MetaMask (version 10.21.2) and capture the parameters when they are invoked. In certain scenarios, users are prompted to sign a data structure via "eth_signTypedData," and we convert it to the relevant transaction parameters according to its content.

In the cat-and-mouse game with anti-phishing systems, phishing websites use various techniques to evade the detection of security organizations [108]. It is the same for TxPhish websites. For example, they maintain a list of IP addresses and Ethereum addresses, which all reach a high visiting frequency. Then they regard those addresses as anti-phishing detectors and block their access. Besides, they pop up useless message windows to hide their phishing behavior. In response, we suggest the following solutions that effectively address these issues.

Setting up a proxy pool. Phishing websites commonly maintain a list of blocked IP addresses that frequently access their platforms. To avoid being identified and added to this list, we've set up a proxy pool with 55 nodes. We randomly select a different proxy every 3 minutes. In addition, IP addresses of these proxies are also updated dynamically.

Setting up an account pool. We've also found that the phishing transaction won't appear when the transaction trigger uses a single Ethereum address for a long time. To circumvent this, we've created a pool of accounts with various token types (including ETH, ERC20, and ERC721 tokens). When visiting phishing websites, we randomly choose an Ethereum address from the pool. We also update the account pool dynamically.

Continuously signing until phishing transactions emerge. Phishing websites often ask users to sign several meaningless messages, such as "Hello," to behave like a legitimate Ethereum project. Moreover, to avoid triggering anti-phishing systems, the phishing transaction won't appear until you sign each message individually. As mentioned in Section 3.2.2, the default settings of MetaMask offer only two types of methods to sign transactions. By contrast, signing useless messages will invoke other methods, like 'personal_sign', which are not applicable for sending transactions. So, in our system, we would also automatically sign these messages continuously after verifying their invoking methods.

4.4 Transaction Detector

Once the transaction parameters are retrieved, the next step is to conduct a simulation execution to determine the transaction's results. For this purpose, we utilize Mopsus [47], an industry-leading transaction pre-execution service provided by BlockSec [20]. Through Mopsus, we are able to obtain a comprehensive overview of the potential profit and loss associated with the transaction.

Detection criteria I: sending or approving tokens to an EOA with nothing in return. As stated in Section 3.2.1, the phishing transaction is designed to withdraw all of the victim's valuable tokens. Since the transaction trigger clicks the "mint", "buy", and "claim" buttons, it means that users should receive new tokens after making a payment. Besides, in a legitimate project on Ethereum, the official account must be an open-source contract address (CA), which allows users to understand how their tokens are processed and trust the project. Therefore, we consider transactions that send or approve tokens to an Externally Owned Account (EOA) but receive nothing in return as phishing transactions. This indicates that the user has fallen victim to a scam.

Detection criteria II: requesting all tokens from accounts with various token types. Nevertheless, the aforementioned detection criteria is inadequate in encompassing all scenarios. For example, there are some TxPhish websites requesting users to send ETH to a customized phishing contract [35]. Since phishing websites are unable to foresee the type of victims' tokens, they will indiscriminately steal all tokens present in the victims' accounts.

Table 2: Evaluation Results of different components in TxPhishScope.

	TP	FP	TN	FN	Precision	Recall
C1 [6]	5634	4758	1679	632	54.2%	89.9%
C2 [7]	5365	4020	2417	901	57.2%	85.6%
C3 [8]	5135	0	6437	1131	100.0%	82.0%
TxPhishScope	4721	0	6437	1545	100.0%	75.3%

As a result, when we switch to an account holding ERC20 tokens or NFTs in these phishing websites, we are prompted to transfer or approve the respective tokens to an account. Therefore, if a website requests all tokens from accounts holding different types of tokens, we also categorize related transactions as phishing transactions.

To save evidence of TxPhish websites, we would store source codes of the website, screenshots, and parameters of the phishing transaction. Additionally, we would leverage *urlscan* [62] to conduct scans on all identified TxPhish websites. This will enable individuals to search the scan results for more information.

4.5 Evaluation of TxPhishScope

To demonstrate the precision and recall metrics of different components in TxPhishScope, we collect relevant datasets and conduct experiments.

TxPhish website dataset. As Chainabuse [25] provides public Ethereum phishing reports [26] that are thoroughly verified by maintainers, we extract TxPhish websites from this valuable resource. We consider any report that includes details about phishing websites and associated accounts as a TxPhish report. Specifically, we collect 6,266 TxPhish websites from reports spanning the period between February 1, 2023, and July 31, 2023.

Benign website dataset. We collect 6,437 official websites of dapps across 10 widely used blockchains from CoinMarketCap [29], DappRadar [34] and CoinGecko [28], which are reputable Web3 data providers. In detail, the benign dataset comprises 2,683 Ethereum projects, 1,606 BNB projects, 502 Polygon projects, 468 Solana projects, 329 Avalanche projects, 285 Arbitrum projects, 211 Fantom projects, 175 TRON projects, 92 Cronos projects, and 86 Klaytn projects.

Evaluation details. We evaluate the Website Scanner and Transaction Trigger & Detector with the above datasets. Since some TxPhish websites have gone offline when reported, we fail to verify whether our tool is capable of detecting them. For simplicity, we assume that TxPhish websites with the same phishing accounts utilize identical toolkits, including the same type of website files, phishing transactions and logic of triggering transactions. For an offline TxPhish report, if the website is not found in our database while the same phishing account exists in our database, we consider the Website Crawler and Transaction Trigger & Detector in our tool are capable of detecting them.

Evaluation results. Table 2 presents the precision and recall ratio of different components in TxPhishScope. Given that a legitimate Web3 project cannot produce phishing transactions as mentioned in Section 4.4, we don't discover any false positives generated by TxPhishScope during the evaluation. Meanwhile, TxPhishScope exhibits a detection rate of **75.3%** for TxPhish websites within the dataset. In addition, as we continually optimize TxPhishScope to identify more TxPhish websites, the inability to detect certain TxPhish websites in the past does not imply a failure to detect them now. Nevertheless, due to the short lifespan of phishing websites, we are unable to reevaluate TxPhishScope using historical data. Therefore, the recall ratio of TxPhishScope in the evaluation may not accurately reflect its true effectiveness. And in Section 4.6, we deploy TxPhishScope for over eight months to demonstrate its efficiency in detecting TxPhish websites.

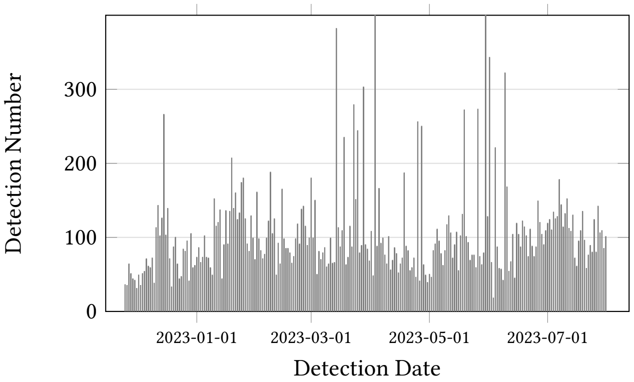


Figure 4: Number of TxPhish Websites Detected by TxPhishScope from November 25, 2022, to July 31, 2023. TxPhishScope identifies around 105.8 TxPhish websites every day.

4.6 Discovering TxPhish Websites in the Wild

To detect TxPhish websites on a large scale and protect users, we have deployed TxPhishScope for over eight months. Figure 4 shows the number of TxPhish websites detected by TxPhishScope. Specifically, from November 25, 2022, to July 31, 2023, we detected **26,333** TxPhish websites and **3,486** phishing accounts. Among them, **93.1%** of TxPhish websites are first discovered by us. To block detected TxPhish websites, prevent users' loss, and verify our results, we organize the evidence and report them to MetaMask [36] and Forta [39] on a daily basis. Besides, we also report these phishing accounts to etherscan. Since MetaMask, etherscan is the most popular wallet and browser on Ethereum, any legitimate projects mistakenly reported by us would contact us quickly. Maintainers of Forta would check our phishing reports, too. Up to now, we haven't received any false positive feedback.

In order to compare our reported TxPhish websites with others, we have collected all phishing websites reported in MetaMask [36] and Chainabuse [26] since February 1, 2023. As some phishing websites have been taken down or are no longer active when reported, we cannot judge whether they are TxPhish websites. So we only compare the number of alive TxPhish websites reported by us and others. To ensure the accuracy of the comparison, we crawl the reported phishing websites every hour for manual checks. For simplicity, we consider any websites that prompt users to connect their wallet as TxPhish websites. Figure 5 summarizes the number

[6] We refer to Domain Scoring Algorithm as C1 for short.

[7] We refer to Website Crawler as C2 for short.

[8] We refer to Transaction Trigger & Detector as C3 for short.

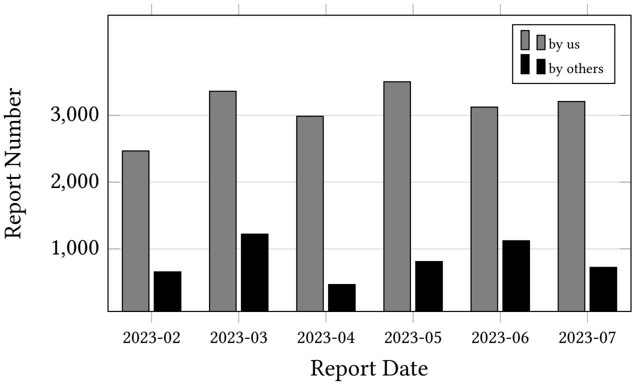


Figure 5: Number of TxPhish Websites First Reported by us and Others from February 1, 2023, to July 31, 2023. 78.9% of TxPhish websites are first reported by us.

of TxPhish websites reported by us and others. Statistically, from February 1, 2023, to July 31, 2023, we first reported **18,652** TxPhish websites, making up **78.9%** of all TxPhish websites. In summary, TxPhishScope is the largest TxPhish website detection system.

In addition to effectively blocking a significant number of TxPhish websites, we also helped the community in other ways. For instance, we presented incriminating evidence for **four** phishing accounts, along with their fund flow of **$1.5 million**, to assist in the restitution of funds to victims. Meanwhile, we discovered bugs for **six** Ethereum projects and received their appreciation. Specifically, the project party fails to check whether users set the right blockchain, users would lose money if they don't check the transaction carefully.

4.7 Comparison with Existing Works

Research efforts aimed at detecting phishing websites have been in continuous development for over a decade. Generally, the detection methods can be implemented based on features extracted from URL string [75, 84, 92, 94, 100], web page components [88, 93, 102], visual elements [71, 72, 76, 81, 89, 90]. In the Domain Scoring Algorithm of TxPhishScope, we draw inspiration from them and incorporate techniques, such as editing distance [75, 84, 92, 94, 100] and keyword matching [74, 99], to discover potentially suspicious websites. The key observation of these works is that phishing websites typically utilize URLs and web page designs that closely resemble those of legitimate websites to mislead and deceive users.

In contrast, the primary method utilized to detect TxPhish websites is through phishing transactions, which also enables the efficient extraction of phishing accounts. We design TxPhishScope to achieve a large-scale detection of TxPhish websites and phishing accounts. In statistics, from February 1, 2023, to July 31, 2023, we discovered **26,333** TxPhish websites and **3,486** phishing accounts with no false positives.

4.8 Limitations

Once the implementation details of TxPhishScope are understood, it may be possible for individuals to devise targeted techniques to evade our detection mechanisms. More specifically, TxPhish websites can attempt bypassing the detection of TxPhishScope during various stages, such as when scoring domains, initiating phishing transactions, or classifying phishing transactions. Nevertheless, due to our modular design, TxPhishScope is extensible to address potential issues that may arise in the future.

While the design of TxPhishScope is straightforward, it has proven highly effective in detecting numerous TxPhish websites, making it a valuable data source for measurement purposes. Next, we will conduct an in-depth analysis of TxPhish websites in Section 5 and examine the fund flows associated with phishing accounts in Section 6. These analyses will provide valuable insights into the nature and impact of TxPhish activities on Ethereum.

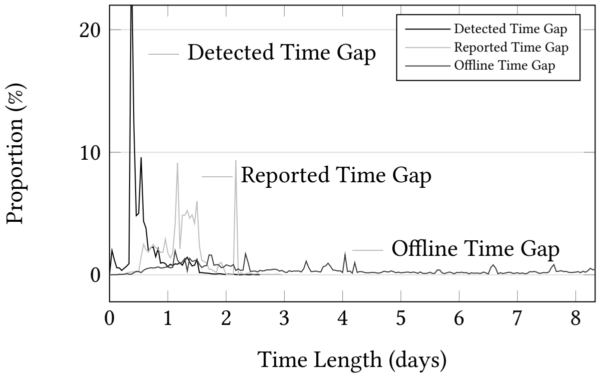


Figure 6: Distribution of Detected, Reported, and Offline Time Gaps for TxPhish Websites. The time gap is calculated from certificate registration. TxPhish websites remain active for 113 hours on average. We can detect and report most of (more than 86.2%) them within one day and two days [9], respectively.

5 ANALYSIS OF TXPHISH WEBSITES

In order to gain a more comprehensive understanding of TxPhish websites and provide guidance for anti-phishing efforts, we conduct a thorough analysis of the phishing websites detected by TxPhishScope and other reporters mentioned in Section 4.6. As a result, we identify three key properties of these phishing websites: short lifespan, low cost, and fast update speed.

5.1 Short Lifespan

From February 1, 2023, to July 31, 2023, we successfully gathered the lifespan data of **12,327** TxPhish websites detected by TxPhishScope. Our data collection includes the time of certificate registration, detection, reporting, and shutdown of these websites. As can be seen from Figure 6, since certificate registration, we detect and report TxPhish websites in **15** hours and **31** hours, on average. Once detected, these websites remain online for an average of **113** hours. Furthermore, for **69.4%** of TxPhish websites, their lifespans from certificate issuance to shutdown are within **seven** days. These

[9] There may be some time latency when we adjust and optimize TxPhishScope.

Table 3: Top-5 Certificate Registrars of TxPhish Websites. 83.3% of TxPhish websites use free certificates to lower Costs.

Certificate Registrar	Number	Proportion
Let's Encrypt	12,725	67.4%
Google Trust Services	2,634	13.9%
Sectigo Limited	1,751	9.3%
cPanel	1,031	5.5%
ZeroSSL	364	1.9%

findings underscore the significance of timely reporting and blocking of TxPhish websites, enabling swift protection for the majority of users. Moreover, our system demonstrates remarkable efficiency, successfully detecting and reporting over **86.2%** of these websites within **one** day and **two** days, respectively.

Despite the short lifespan of TxPhish websites, criminal activities carried out by associated phishing groups persist. Typically, when a TxPhish is about to go offline, the same phishing account is likely to be present on another website quickly. Based on our detection results, we have identified **445** active phishing accounts that have been in operation for over a month. Taking inspiration from this observation, we proceeded to enhance TxPhishScope by incorporating additional detection criteria. Specifically, if the target address matches a known phishing entry in our database, we swiftly identify the suspicious website as a TxPhish website.

5.2 Low Cost

Since phishing syndicates tend to launch large quantities of phishing websites at the same time, they attempt to minimize their cost. To investigate this trend, we gather WHOIS records and certificate information for **18,884** TxPhish websites. Disregarding the expenses associated with coding, we analyze the phishing cost from the perspective of domain and certificate registration.

Registering certificates for free. As can be seen from Table 3, **83.3%** of TxPhish websites use free certificates offered by Let's Encrypt [43], Google Trust Services [41], and ZeroSSL [69].

Using a shared parent domain to save expense. In an effort to reduce domain registration costs, numerous phishing websites share a common parent domain. For instance, we come across **195** TxPhish websites that share the parent domain *whitelist-airdrops.com*, which was registered through *WEBCC* [65]. And all of the certificates are issued by *Let's Encrypt* freely. According to the price list of *WEBCC* [66], the registration fee for *whitelist-airdrops.com* for a year is $24.6, implying that the registration fee for both the domain and the certificate of **.whitelist-airdrops.com* is only **$0.13**. Additionally, out of the **18,884** TxPhish websites, we have found **6,754** (**35.8%**) websites that share **673** parent domains. Driven by this phenomenon, we take action against these TxPhish websites that share parent domains by directly reporting and blocking their parent domains.

5.3 Fast Update Speed

In the world of Web3, new projects and technologies are continuously emerging. However, any novel development on Ethereum can be leveraged by scammers to deceive and cheat unsuspecting users. As a result, different from traditional phishing websites, TxPhish websites update rapidly.

Table 4: Popular Projects and Related TxPhish Websites around their Trending Phases

Popular Projects	Most Trending Phases	Number of Related TxPhish Websites
TrustPad [61]	January 2023	176 in 22 days
Blur [21]	February 2023	265 in 34 days
zkSync [70]	March 2023	168 in 20 days
Arbitrum [16]	March 2023	641 in 24 days
AIDOGE [12]	April 2023	104 in 18 days
pepe [50]	May 2023	434 in 38 days

The content of TxPhish websites updates rapidly. On Ethereum, hot events happen every day, making associated projects get popular quickly. Exploiting this, scammers deploy large numbers of similar phishing websites at the same time. Table 4 shows the number of related TxPhish websites for several popular projects in 2023. For instance, during the Arbitrum [16] airdrop around March 2023, which attracted many users, we identified **641** fake Arbitrum TxPhish websites in **24** days. Based on this observation, our crawlers are continuously gathering the most recent information on popular Web3 projects from CoinMarketCap [29], DappRadar [34], and CoinGecko [28].

The type of phishing transactions updates rapidly. In Section 3.2.1, we summarize three types of phishing transactions. Nevertheless, with new technologies appearing, scammers might leverage them to develop novel forms of phishing transactions. As a result, the toolkits of TxPhish websites are updated every few weeks [30, 42]. For instance, the zero-dollar purchase phishing technique was previously limited to Opensea orders. However, after observing a significant increase in the number of Blur users, we also discover evidence of zero-dollar purchase phishing for Blur orders [68] on TxPhish websites. Through file hash matching, we identify **270** similar TxPhish websites in March 2023. While we have provided a comprehensive overview of phishing transactions, there is always the possibility of new types emerging in the future. Additionally, we will diligently monitor TxPhish toolkits to stay up-to-date with the latest information on phishing transactions.

5.4 Comparison with Existing Works

Previous research has measured the life cycle and ecosystem of phishing websites from various perspectives. Oest et al. [96] present the study of anti-phishing techniques by analyzing phishing toolkits. Oest et al. [97] uncover the end-to-end life cycle and effectiveness of phishing attacks. PhishTime [95] measures the effectiveness of anti-phishing blacklists. CrawlPhish [107] conducts a large-scale analysis of cloaking techniques deployed on the client side. Bijmans et al. [74] measure the Dutch phishing campaigns from the perspectives of phishing kits, domains, and phishing website deployments. Kondracki et al. [85] depict the *Man-in-the-Middle* (*MITM*) phishing toolkits, forwarding requests between the victims and target servers. And they propose a detection system based on network timing and TLS library features. Zhang et al. [109] summarize all anti-phishing techniques and implement an automated client-side framework to assist users in evading phishing websites. The framework achieves this by making legitimate users appear like anti-phishing crawlers.

Although our analysis of lifecycle for TxPhish websites is partially overlapped with existing works [74, 87], we measure their lifespan in a more comprehensive way, including certificate register, detection, report, and offline time. In addition, we reveal the remarkably low cost of TxPhish websites from the perspectives of certificate and domain registration. What's more, we unveil the rapid update speed of TxPhish toolkits, both in terms of webpage content and phishing transaction types.

6 ANALYSIS OF FUND FLOW

To aid in the recovery of funds for victims, it is crucial to conduct a thorough analysis of the fund flow associated with these phishing accounts. To this end, we first extract history transactions from active phishing accounts identified by TxPhishScope. Then we summarize common techniques for handling phishing funds. Next, based on these insights, we propose to build a fund flow graph for each account. Finally, from the results of the fund flow graphs, we present the distribution of phishing fund flows and behavioral patterns associated with each account.

6.1 Collection of Phishing Transactions

From detection results of TxPhishScope in Section 4.6, we have identified active phishing accounts (having over 10 transactions in one year) and their corresponding incoming transactions of funds (amounting to more than $100). Despite the incoming funds not necessarily originating from a victim, it is still under the possession of the phishing group, which can provide valuable insights into the fund flow patterns of the related group. To expand our dataset, we also include the target addresses found in the *TransferFrom* function call for approval phishing. In statistics, we collected **630** phishing accounts and **10,705** incoming transactions. Since we have manually verified these addresses and shared relevant evidence of phishing with etherscan, they can be regarded as reliable and accurate data.

6.2 Techniques for Handling Phishing Funds

Figure 7 illustrates methods for managing funds in a phishing account. More precisely, the process involves either directly exchange tokens through Centralized Exchanges (CEXs) or first laundering money to conceal the illegality of funds.

Exchanging tokens via CEXs. In a typical scam scenario, scammers can convert their illicit tokens into cash or other types of tokens by transferring them to a centralized exchange (CEX). However, if scammers deposit these tokens directly into a CEX, it is relatively easy for the CEX to trace the scamming account back to the actual identity of the scammer in real life. To avoid detection, scammers need to launder these tokens and transfer them to an address that appears legitimate.

Laundering money via zero-knowledge proof-based mixers. In this scenario, the user can deposit tokens into several smart contracts, and subsequently withdraw these deposits to a different address with a cryptographic proof. This makes it impossible to link the deposit to the withdrawal, thereby enabling users to launder their funds without revealing their identity. The most commonly used service for zero-knowledge proof-based money laundering on Ethereum are Tornado Cash [59], Aztec Connect [17], Secret Network [57].

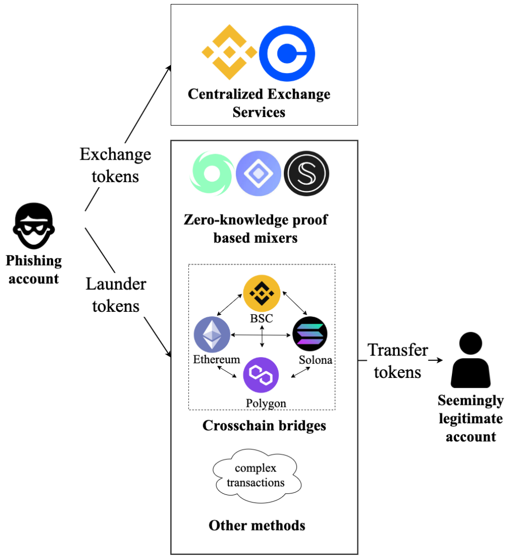


Figure 7: Methods for Managing Funds in a Phishing Account. It can either exchange tokens directly via Centralized Exchanges (CEXs) or first launder money through mixers, cross-chain bridges, or other means.

Laundering money via cross-chain bridges. The purpose of cross-chain bridges is to facilitate asset liquidity among different blockchains. However, since there are numerous cross-chain bridges with varying design mechanisms, it can be challenging to establish an automatic identification scheme of the source and destination addresses in cross-chain transactions. Consequently, scammers resort to launching multiple cross-chain transactions to evade detection.

Laundering money via other methods. Apart from the three methods mentioned above, we have encountered several phishing accounts with intricate fund flow transactions. In such cases, tokens are transferred among dozens or even hundreds of accounts. It's possible that these accounts are involved in the distribution of spoils among a vast phishing group. Also, they may leverage money laundering services provided by underground economy organizations. Nevertheless, due to the absence of off-chain information and labels associated with these accounts, we are unable to determine the fund flow for these cases promptly.

6.3 Constructing Fund Flow Graphs

To clearly demonstrate the path and destination of fund flow, we propose to construct a fund flow graph. For each phishing account, we would analyze its incoming transactions one by one and build its fund flow graph. The construction process is outlined below.

- **S1: extract the sender and recipient from the transaction.** To begin analyzing a transaction, we extract the sender and recipient addresses, as well as the token value transferred.
- **S2: add the fund flow into the graph.** Within the graph, each node represents an account address, and the edge value denotes

Table 5: Distribution of Fund Flow Targets

Type of Fund Flow Targets	Amount	Proportion	Most Popular Targets	Amount	Proportion
Centralized Exchange services	$43.7M	54.0%	Binance	$13.6M	16.8%
			OKX	$8.9M	11.0%
			eXch	$4.1M	5.0%
			SimpleSwap	$2.6M	3.2%
Zero-knowledge proof-based mixers	$9.5M	11.8%	Tornado Cash	$9.3M	11.6%
			Secret Network	$0.2M	0.2%
Cross-chain bridges	$2.4M	2.9%	TransitSwap	$0.9M	1.1%
			SWFT	$0.6M	0.7%
			Socket	$0.2M	0.2%
Remain in accounts	$7.0M	8.7%	0x6345****060e [10]	$0.9M	1.1%
			0xd6b8****9287 [11]	$0.4M	0.5%
			0x2992****5479 [12]	$0.3M	0.4%
Other methods	$18.2M	22.6%	Stake	$2.7M	3.4%

the list of transactions made between two addresses. If a node or edge is absent from the graph, we will add them individually.

- **S3: determine whether or not to terminate.** Sometimes, if the money laundering process is too complex, our analysis could get stuck in an endless loop. To prevent this, we have established three conditions for terminating the analysis. First, if *the recipient address is a CEX service provider*, we can halt and save related evidence as it can assist victims in recovering their funds. Secondly, if *the recipient address is a mixer or cross-chain bridge*, we also terminate the analysis due to the difficulty of analysis. Thirdly, if *the recursive depth surpasses 10* or *the recipient address is a super node having more than 3000 transactions*, we will stop as the further fund flow is too complex. For other situations, we will progress to **S4**.
- **S4: search for following transfer transactions and go to S1.** We will track the fund flow of the token extracted in **S1** for the subsequent transfers of the recipient address. If there are any following transfers associated with the token, we would go to **S1** and track the same amount of tokens again. Any remaining tokens in the recipient address will be labeled as "remained". To avoid analyzing the same amount of tokens multiple times, We would keep track of the number of tokens already analyzed in each transfer.

6.4 Distributions of Fund Flow targets

Table 5 shows the distribution of fund flow targets. As the value of cryptocurrency fluctuates on a daily basis, we monitored a total inflow of **$80.8 million** worth of tokens into various targets from the incoming tokens valued at **$81.0 million**. Among them, we could successfully trace **80.8%** of funds to public projects or accounts on Ethereum.

$43.7 million flowed into CEXs. Unlike Ethereum attacks that cause losses of millions of dollars across several transactions, phishing scams typically only earn a few thousand dollars from a single account. In statistics, among transactions of all phishing accounts, there are **8364** (**78.1%**) transfers earning less than $5000. Due to the difficulty for victims to gather their evidence and contact authorities, phishing gangs feel emboldened to transfer their illicit funds to centralized exchanges (CEXs) and exchange them for cash or other types of tokens. As a result, an astonishing sum of **$43.7 million** flowed into CEXs, with Binance, OKX, eXch, and SimpleSwap emerging as the top four, accounting for **16.8%**, **11.0%**, **5.0%**, and **3.2%** of the total funds respectively. By providing sufficient criminal evidence of phishing accounts and their associated fund flows to CEXs, victims can rapidly identify the true identity of the scammers.

$9.5 million flowed into mixers. Since the privacy of funds flowing into mixers is ensured by zero-knowledge proof, It is impossible to map the deposit to the withdrawal. This makes mixers a more desirable option for money laundering compared to cross-chain bridges. Among the three commonly used mixers, Tornado Cash is the most favored, receiving approximately **11.6%** of the funds. Although Tornado Cash has been sanctioned by U.S. Treasury since August 8, 2022 [63], scammers still prefer it as their primary option for laundering money.

$7.0 million remained in accounts. In order to conceal the final destination of their illegal tokens, certain scammers opt to delay processing phishing funds for a specified period. As of August 10, 2023, approximately **8.7%** of these funds tend to linger in particular accounts. Since we only tracked phishing transactions mentioned above, there may be discrepancies between the actual token values held in an account and our estimated values. For instance, according to the fund flow graph, 0xd6b8****9287 retains around $0.402 million, whereas a manual inspection of the account reveals that it holds tokens worth $0.441 million. The left $0.040 million could be attributed to other sources, but it's also plausible that it constitutes illegal funds obtained by the phishing groups involved. We strongly recommend that victims closely monitor the flow of funds in these accounts and take immediate action if any transfers are made to cryptocurrency exchanges (CEXs) or other public services.

$18.2M was processed via other methods. Out of all the fund flow tokens, a total of **$18.2 million** was not processed with the three methods mentioned above. Despite this, it has come to our attention that **$2.7 million (3.4%)** was transferred into stake.com [58], an online casino. It is also a form of money laundering, since we fail to trace the phishing funds to a specific transaction, especially when the scammers exhaust the victims' tokens.

[10] 0x6345285f2f9ecbd2ff75605c6b5fd2c08436060e

[11] 0xd6b8c482c1f06265223b8bd0ad6efbf1214a9287

[12] 0x29923892bea33877dcddd3ea321b2c086a735479

Table 6: Fund Flow Behavioral Patterns of Phishing Accounts. 44.8% of phishing accounts would exchange most of (more than 80%) their funds. While 13.2% of them would first launder money via public services.

Fund Flow Pattern	Number of Phishing Accounts
Exchanging tokens via CEXs	282 (44.8%)
Launder money via public services	83 (13.2%)
Above 2 ways combined	83 (13.2%)
Remain in accounts	35 (5.6%)
Uncertain	147 (23.3%)

Meanwhile, we noticed that **$6.7 million (8.2%)** flowed into **280** supernodes having more than 3000 transactions. However, due to insufficient on-chain data, we are unable to determine their specific purposes. These accounts may be used for public services, but they could also belong to black market entities involved in money laundering activities.

57.9% of phishing accounts have fixed fund flow behavioral patterns. As can be seen from Table 6, **44.8%** of phishing accounts exchange most of their funds via CEXs. And **13.2%** of them first launder most of their money through public services (mixers, cross-chain bridges, Stake, and payment services). This indicates that they possess fixed fund flow behavioral patterns. While **13.2%** of phishing accounts process their funds both via CEXs and public money laundering services, this may be attributed to the sharing of profits among members of the phishing group. In conclusion, the key to enforcing sanctions against scammers and facilitating the recovery of funds for victims lies in the collaborative efforts of public service providers, law enforcement, and the victims themselves.

6.5 Comparison with Existing Works

In the last few years, research works on detecting and measuring phishing accounts on Ethereum have flourished. Since the transaction behavior of phishing accounts are different from those legitimate ones, several works have been proposed to extract features from transaction networks and train AI models for identification [77, 78, 83, 86, 101, 103, 105, 106]. Phillips et al. [98] present the distributions of phishing fund source and destination.

To the best of our knowledge, we conduct the most extensive analysis of phishing fund flows on Ethereum, encompassing a vast number of transactions (exceeding **10,000**) and a substantial fund volume (over **$80.0 million**). Our study reveal that a flow of funds exceeding **$43.7** million directed towards CEXs and specific public service accounts, thereby facilitating the recovery of phishing funds. Additionally, we discover that an amount around **$7.0** million still resides within certain phishing accounts, enabling the possibility of monitoring the subsequent fund flow destinations. Moreover, we categorize phishing accounts based on behavioral patterns and identify that **57.9%** of phishing accounts exhibit consistent and fixed fund flow behaviors.

6.6 Limitations

Due to insufficient off-chain information, we are unable to conduct a comprehensive analysis of it. Before the funds eventually make their way to CEXs, mixers, cross-chain bridges, or other money laundering services, a series of intricate transactions take place. The reasons behind these transactions, such as profit distribution or money laundering in illicit market activities, are challenging to elucidate.

7 RELATED WORK

7.1 Detecting Web3 Scams

Several techniques have been proposed to detect Web3 scams. In phishing account detection, since they are different from legitimate ones in transaction frequency and targets, several works have been proposed to extract features from transaction networks and train AI models for identification [78, 86]. For scam token detection, Xia et al. [104] suggest extracting time-series, transaction, investor, and uniswap-specific features to train a classifier. SCSGuard [82] utilizes n-gram features and attention neural network to detect scams in smart contracts from the bytecode of contracts. SADPonzi [79] proposes detecting Ponzi scams in smart contracts by extracting semantic information through symbolic execution and comparing it to summarized Ponzi scheme patterns.

7.2 Measuring Security Issues on Ethereum

The world of Ethereum has faced many security issues since its inception. At the same time, various research endeavors have been undertaken to quantify and comprehend these security concerns. Das et al. [80] present a systematic overview of the NFT ecosystem and uncover associated security issues. Lyu et al. [91] perform a detailed analysis of private transactions and their security implications on Ethereum. Roy et al. [99] study NFT promotion phishing scams on twitter. Then, they develop ML-based models to detect NFT-based phishing websites and fradulent NFT phishing projects on twitter. TxPhish websites that impersonate NFT projects can also be classified as NFT-based phishing websites. Li et al. [87] study the giveaway scams, where users send tokens to a designated address with the expectation of receiving double the amount in return, only to end up receiving nothing in the end. In our research, we concentrate on the large-scale detection of TxPhish websites via phishing transactions, allowing us to conduct a comprehensive measurement of both TxPhish websites and the phishing fund flow.

8 FUTURE WORK

Detecting phishing contracts. As previously stated, numerous phishing contracts have been found to withdraw victims' tokens directly. These contracts often have misleading withdraw function names, such as "Claim Token" or "Security Update", which can be highly deceptive. Currently, the detection of such contracts relies mostly on manual inspection. Automated detection after they have been deployed on-chain is an area for future research.

Analyzing the workflow of TxPhish groups. Due to the potential for TxPhish to generate significant profits in a short period, the number of individuals joining phishing groups has been steadily increasing. Additionally, we have observed numerous websites and Telegram groups selling toolkits specifically designed for TxPhish. Despite these trends, there remains a lack of comprehensive research regarding the advertising and profit distribution processes utilized by phishing groups.

9 CONCLUSION

In this paper, we provide a comprehensive analysis of TxPhish. Then we review existing sources of reports on TxPhish and identify several limitations. To address these issues simultaneously, we introduce a system for detecting TxPhish websites called TxPhishScope. The results of our system show that it is the largest TxPhish website detection system available. Moreover, we provided criminal evidence of **four** phishing accounts with a total fund flow of **$1.5 million**, which could aid in the recovery of funds for victims. Additionally, we discovered bugs in **six** Ethereum projects, avoiding potential user losses and receiving appreciation. Based on the detection results of TxPhishScope, we conduct comprehensive measurements of TxPhish websites and phishing accounts. We reveal the ecosystem of TxPhish websites through their lifespan, cost, and update frequency. Besides, we conduct an in-depth analysis of the fund flow targets and behavioral patterns of phishing accounts. Our discoveries can act as a reference for Ethereum service providers to protect their users from phishing scams and assist in the recovery of victims' funds.

10 ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their comments that greatly helped improve the presentation of this paper. Additionally, the first author of this paper would like to thank Xiaohui Hu and Qiang Liu personally for their help. This work is partially supported by the National Key R&D Program of China (No. 2022YFE0113200) and the National Natural Science Foundation of China (NSFC) under Grant No. U21A20464, No.62172360, No.U21A20467. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of funding agencies.

REFERENCES

[1] 2015. *ERC-20: Token Standard.* Retrieved February 26, 2023 from https://eips.ethereum.org/EIPS/eip-20
[2] 2018. *ERC-1155: Multi Token Standard.* Retrieved February 26, 2023 from https://eips.ethereum.org/EIPS/eip-1155
[3] 2018. *ERC-721: Non-Fungible Token Standard.* Retrieved February 26, 2023 from https://eips.ethereum.org/EIPS/eip-721
[4] 2022. *Be Wary of the TransferFrom Zero Transfer Scam.* Retrieved Mar 1, 2023 from https://medium.com/@slowmist/slowmist-be-wary-of-the-transferfrom-zero-transfer-scam-c64ba0e3bc4d
[5] 2022. *Ethereum Wallet MetaMask Passes 30M Users, Plans DAO and Token.* Retrieved Mar 6, 2023 from https://decrypt.co/95039/metamask-consensys-30-million-users
[6] 2022. *Fake Airdrops, Fake Wallets and Now Fake Exchange Apps.* Retrieved Mar 1, 2023 from https://slowmist.medium.com/fake-airdrops-fake-wallets-and-now-fake-exchange-apps-ddb78f770f73
[7] 2022. *An in-depth look into the infrastructure supporting the "fake wallet" phishing industry.* Retrieved Mar 1, 2023 from https://slowmist.medium.com/an-in-depth-look-into-the-infrastructure-supporting-the-fake-wallet-phishing-industry-79030560f38
[8] 2022. *OpenSea Phishing Attack Led To Loss Of NFTs Worth 641 Ethereum.* Retrieved Mar 6, 2023 from https://www.lowyat.net/2022/266445/opensea-phishing-attack-loss-of-nfts-641-ethereum
[9] 2022. *Uniswap User Loses $8M Worth of Ether in Phishing Attack.* Retrieved Mar 6, 2023 from https://www.coindesk.com/tech/2022/07/12/uniswap-user-loses-8m-worth-of-ether-in-phishing-attack
[10] 2023. *$10,000,000 scammer.* Retrieved May 1, 2023 from https://twitter.com/MetaSleuth/status/1643901208116224000
[11] 2023. *2022 Annual Blockchain Security and AML Analysis Annual Report.* Retrieved Mar 1, 2023 from https://www.slowmist.com/report/2022-Blockchain-Security-and-AML-Analysis-Annual-Report(EN).pdf
[12] 2023. *AIDOGE.* Retrieved Aug 10, 2023 from https://arbdoge.ai/
[13] 2023. *Alchemy.* Retrieved Mar 6, 2023 from https://www.alchemy.com
[14] 2023. *Amazon NFTs, Losing $2m in a phishing attack, $105m payday, Is Bitcoin the best performing asset in the world this year?* Retrieved Mar 6, 2023 from https://fomofix.substack.com/p/amazon-nfts-losing-2m-in-a-phishing?utm_source=twitter&utm_campaign=auto_share&r=1duf4y
[15] 2023. *Approval Phishing stealed 70 WBTC.* Retrieved May 1, 2023 from https://twitter.com/MetaSleuth/status/1638812482021228544
[16] 2023. *Arbitrum.* Retrieved April 10, 2023 from https://arbitrum.io
[17] 2023. *Aztec Connect.* Retrieved April 19, 2023 from https://aztec.network/connect
[18] 2023. *Binance.* Retrieved Mar 6, 2023 from https://www.binance.com/en
[19] 2023. *BLOCKMAGE.* Retrieved April 5, 2023 from https://twitter.com/BlockMageSec
[20] 2023. *BlockSec.* Retrieved April 5, 2023 from https://blocksec.com
[21] 2023. *Blur.* Retrieved Mar 6, 2023 from https://blur.io
[22] 2023. *Certificate Transparency.* Retrieved Mar 28, 2023 from https://certificate.transparency.dev
[23] 2023. *CertikAlert.* Retrieved Mar 6, 2023 from https://twitter.com/CertiKAlert
[24] 2023. *certstream-go.* Retrieved Mar 28, 2023 from https://github.com/CaliDog/certstream-go.git
[25] 2023. *Chainabuse.* Retrieved Mar 1, 2023 from https://www.chainabuse.com
[26] 2023. *Chainabuse Ethereum Phishing Scam Reports.* Retrieved Mar 1, 2023 from https://www.chainabuse.com/category/phishing?page=0&filter=ETH
[27] 2023. *Coinbase.* Retrieved Mar 6, 2023 from https://www.coinbase.com
[28] 2023. *CoinGecko.* Retrieved Aug 2, 2023 from https://www.coingecko.com//
[29] 2023. *CoinMarketCap.* Retrieved Aug 2, 2023 from https://coinmarketcap.com/
[30] 2023. *Crypto Drainers | Multichain Drainer.* Retrieved July 26, 2023 from https://t.me/ethdrainer
[31] 2023. *Crypto Google search ad phishing has resulted $4.16 million loss.* Retrieved April 29, 2023 from https://twitter.com/WuBlockchain/status/1651514902408986626
[32] 2023. *Crypto phishing scammer Monkey Drainer shuts down services.* Retrieved Mar 6, 2023 from https://cryptoslate.com/crypto-phishing-scammer-monkey-drainer-shuts-down-services
[33] 2023. *Cyber Security Firm CertiK Unmasks Scammers Linked To $4.3M Porsche NFT Phishing Scam.* Retrieved Mar 6, 2023 from https://www.business2community.com/nft-news/cyber-security-firm-certik-unmasks-scammers-linked-to-4-3m-porsche-nft-phishing-scam-02617446
[34] 2023. *DappRadar.* Retrieved Aug 2, 2023 from https://dappradar.com/
[35] 2023. *Demystifying Profit Sharing in Inferno Drainer.* Retrieved July 19, 2023 from https://blocksecteam.medium.com/demystifying-profit-sharing-in-inferno-drainer-2e8a9afb974b
[36] 2023. *eth-phishing-detect.* Retrieved Mar 1, 2023 from https://github.com/MetaMask/eth-phishing-detect/pulls
[37] 2023. *etherscan.* Retrieved April 1, 2023 from https://etherscan.io
[38] 2023. *Fake_Phishing8210 on etherscan.com has taken $1.24M in USDC from a victim.* Retrieved Mar 6, 2023 from https://twitter.com/CertiKAlert/status/1623131855661805571
[39] 2023. *Forta.* Retrieved Jul 26, 2023 from https://forta.org/
[40] 2023. *Forta Scam Detector.* Retrieved Jul 27, 2023 from https://docs.forta.network/en/latest/scam-detector-bot/
[41] 2023. *Google Trust Services.* Retrieved April 9, 2023 from https://pki.goog
[42] 2023. *Inferno Multichain Drainer.* Retrieved July 26, 2023 from https://t.me/Inferno_Drainer
[43] 2023. *Let's Encrypt.* Retrieved April 9, 2023 from https://letsencrypt.org
[44] 2023. *MetaMask.* Retrieved Mar 1, 2023 from https://metamask.io
[45] 2023. *MetaMask disables eth_sign by default.* Retrieved Aug 11, 2023 from https://github.com/MetaMask/metamask-mobile/issues/5676
[46] 2023. *MetaMask Docs.* Retrieved Mar 18, 2023 from https://docs.metamask.io
[47] 2023. *Mopsus.* Retrieved April 5, 2023 from https://mopsus.blocksec.com
[48] 2023. *Moralis.* Retrieved Mar 6, 2023 from https://moralis.io
[49] 2023. *OpenSea.* Retrieved Mar 6, 2023 from https://opensea.io
[50] 2023. *pepe.* Retrieved Jul 26, 2023 from https://www.pepe.vip
[51] 2023. *Phishing Activity Trends Report: 2nd Quarter, 2021.* Retrieved Mar 26, 2023 from https://docs.apwg.org/reports/apwg_trends_report_q2_2021.pdf
[52] 2023. *Phishing Reports Collected from Twitter.* Retrieved Mar 2, 2023 from https://github.com/MetaMask/eth-phishing-detect/pull/11786
[53] 2023. *Phishing Reports Collected from urlscan.* Retrieved Mar 2, 2023 from https://github.com/MetaMask/eth-phishing-detect/pull/11736
[54] 2023. *Phishing Victim Reports.* Retrieved Mar 2, 2023 from https://github.com/MetaMask/eth-phishing-detect/issues/11742
[55] 2023. *Puppeteer.* Retrieved April 3, 2023 from https://pptr.dev
[56] 2023. *Robinhood and NFT project Azukis Twitter hacked; 122 NFTs worth 484.99 ETH stolen from the latter.* Retrieved Mar 6, 2023 from https://gemhodlers.com/robinhood-and-nft-project-azukis-twitter-hacked-122-nfts-worth-484-99-eth-stolen-from-the-latter
[57] 2023. *Secret Network.* Retrieved April 19, 2023 from https://scrt.network
[58] 2023. *Stake: Crypto Casino & Sports Betting - BTC Casino Online.* Retrieved April 26, 2023 from https://stake.com

[59] 2023. *Tornado Cash.* Retrieved April 14, 2023 from https://ipfs.io/ipns/tornadocash.eth

[60] 2023. *Total Value Locked All Chains.* Retrieved April 30, 2023 from https://defillama.com/chains

[61] 2023. *TrustPad.* Retrieved April 10, 2023 from https://trustpad.io

[62] 2023. *urlscan.* Retrieved Mar 2, 2023 from https://urlscan.io

[63] 2023. *U.S. Treasury Sanctions Notorious Virtual Currency Mixer Tornado Cash.* Retrieved April 26, 2023 from https://home.treasury.gov/news/press-releases/jy0916

[64] 2023. *Venom Drainer has Drained $27M from 15k victims.* Retrieved April 29, 2023 from https://twitter.com/realScamSniffer/status/1642813130454765568

[65] 2023. *WEBCC.* Retrieved April 12, 2023 from https://www.web.cc

[66] 2023. *WEBCC domain register price list.* Retrieved April 12, 2023 from https://www.web.cc/pricing.php

[67] 2023. *What is Ethereum.* Retrieved February 26, 2023 from https://ethereum.org/en/what-is-ethereum

[68] 2023. *Zero dollar purchase phishing in Blur.* Retrieved April 10, 2023 from https://twitter.com/MetaSleuth/status/1633318417938939905

[69] 2023. *ZeroSSL.* Retrieved April 9, 2023 from https://zerossl.com

[70] 2023. *zkSync.* Retrieved April 10, 2023 from https://zksync.io

[71] Sahar Abdelnabi, Katharina Krombholz, and Mario Fritz. 2020. Visualphishnet: Zero-day phishing website detection by visual similarity. In *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security.* 1681–1698.

[72] Sadia Afroz and Rachel Greenstadt. 2011. Phishzoo: Detecting phishing websites by looking at them. In *2011 IEEE fifth international conference on semantic computing.* IEEE, 368–375.

[73] Andreas M Antonopoulos and Gavin Wood. 2018. *Mastering ethereum: building smart contracts and dapps.* O'reilly Media.

[74] Hugo LJ Bijmans, Tim M Booij, Anneke Schwedersky, Aria Nedgabat, and Rolf van Wegberg. 2021. Catching Phishers By Their Bait: Investigating the Dutch Phishing Landscape through Phishing Kit Detection.. In *USENIX Security Symposium.* 3757–3774.

[75] Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. 2010. Lexical feature based phishing URL detection using online learning. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security.* 54–60.

[76] Ahmet Selman Bozkir and Murat Aydos. 2020. LogoSENSE: A companion HOG based logo detection scheme for phishing web page and E-mail brand recognition. *Computers & Security* 95 (2020), 101855.

[77] Liang Chen, Jiaying Peng, Yang Liu, Jintang Li, Fenfang Xie, and Zibin Zheng. 2020. Phishing scams detection in ethereum transaction network. *ACM Transactions on Internet Technology (TOIT)* 21, 1 (2020), 1–16.

[78] Weili Chen, Xiongfeng Guo, Zhiguang Chen, Zibin Zheng, and Yutong Lu. 2020. Phishing Scam Detection on Ethereum: Towards Financial Security for Blockchain Ecosystem.. In *IJCAI*, Vol. 7. 4456–4462.

[79] Weimin Chen, Xinran Li, Yuting Sui, Ningyu He, Haoyu Wang, Lei Wu, and Xiapu Luo. 2021. Sadponzi: Detecting and characterizing ponzi schemes in ethereum smart contracts. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 5, 2 (2021), 1–30.

[80] Dipanjan Das, Priyanka Bose, Nicola Ruaro, Christopher Kruegel, and Giovanni Vigna. 2022. Understanding security issues in the NFT ecosystem. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security.* 667–681.

[81] Anthony Y Fu, Liu Wenyin, and Xiaotie Deng. 2006. Detecting phishing web pages with visual similarity assessment based on earth mover's distance (EMD). *IEEE transactions on dependable and secure computing* 3, 4 (2006), 301–311.

[82] Huiwen Hu, Qianlan Bai, and Yuedong Xu. 2022. Scsguard: Deep scam detection for ethereum smart contracts. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).* IEEE, 1–6.

[83] Arkan Hammoodi Hasan Kabla, Mohammed Anbar, Selvakumar Manickam, and Shankar Karupayah. 2022. Eth-PSD: A machine learning-based phishing scam detection approach in ethereum. *IEEE Access* 10 (2022), 118043–118057.

[84] Taeri Kim, Noseong Park, Jiwon Hong, and Sang-Wook Kim. 2022. Phishing URL Detection: A Network-based Approach Robust to Evasion. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security.* 1769–1782.

[85] Brian Kondracki, Babak Amin Azad, Oleksii Starov, and Nick Nikiforakis. 2021. Catching transparent phish: analyzing and detecting MITM phishing toolkits. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security.* 36–50.

[86] Sijia Li, Gaopeng Gou, Chang Liu, Chengshang Hou, Zhenzhen Li, and Gang Xiong. 2022. TTAGN: Temporal transaction aggregation graph network for ethereum phishing scams detection. In *Proceedings of the ACM Web Conference 2022.* 661–669.

[87] Xigao Li, Anurag Yepuri, and Nick Nikiforakis. 2023. Double and Nothing: Understanding and Detecting Cryptocurrency Giveaway Scams. In *Network and Distributed Systems Security (NDSS) Symposium.*

[88] Yukun Li, Zhenguo Yang, Xu Chen, Huaping Yuan, and Wenyin Liu. 2019. A stacking model using URL and HTML features for phishing webpage detection. *Future Generation Computer Systems* 94 (2019), 27–39.

[89] Yun Lin, Ruofan Liu, Dinil Mon Divakaran, Jun Yang Ng, Qing Zhou Chan, Yiwen Lu, Yuxuan Si, Fan Zhang, and Jin Song Dong. 2021. Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages.. In *USENIX Security Symposium.* 3793–3810.

[90] Ruofan Liu, Yun Lin, Xianglin Yang, Siang Hwee Ng, Dinil Mon Divakaran, and Jin Song Dong. 2022. Inferring phishing intention via webpage appearance and dynamics: A deep vision based approach. In *31st USENIX Security Symposium (USENIX Security 22).* 1633–1650.

[91] Xingyu Lyu, Mengya Zhang, Xiaokuan Zhang, Jianyu Niu, Yinqian Zhang, and Zhiqiang Lin. 2022. An Empirical Study on Ethereum Private Transactions and the Security Implications. *arXiv preprint arXiv:2208.02858* (2022).

[92] Justin Ma, Lawrence K Saul, Stefan Savage, and Geoffrey M Voelker. 2009. Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.* 1245–1254.

[93] Anutthamaa Martin, Na Anutthamaa, M Sathyavathy, Marie Manjari Saint Francois, Dr V Prasanna Venkatesan, et al. 2011. A framework for predicting phishing websites using neural networks. *arXiv preprint arXiv:1109.1074* (2011).

[94] Rami M Mohammad, Fadi Thabtah, and Lee McCluskey. 2012. An assessment of features related to phishing websites using an automated technique. In *2012 international conference for internet technology and secured transactions.* IEEE, 492–497.

[95] Adam Oest, Yeganeh Safaei, Penghui Zhang, Brad Wardman, Kevin Tyers, Yan Shoshitaishvili, Adam Doupé, and Gail-Joon Ahn. 2020. Phishtime: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists. In *Proceedings of the 29th USENIX Conference on Security Symposium.* 379–396.

[96] Adam Oest, Yeganeh Safei, Adam Doupé, Gail-Joon Ahn, Brad Wardman, and Gary Warner. 2018. Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis. In *2018 APWG Symposium on Electronic Crime Research (eCrime).* IEEE, 1–12.

[97] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. 2020. Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. In *29th {USENIX} Security Symposium ({USENIX} Security 20).*

[98] Ross Phillips and Heidi Wilder. 2020. Tracing cryptocurrency scams: Clustering replicated advance-fee and phishing websites. In *2020 IEEE international conference on blockchain and cryptocurrency (ICBC).* IEEE, 1–8.

[99] Sayak Saha Roy, Dipanjan Das, Priyanka Bose, Christopher Kruegel, Giovanni Vigna, and Shirin Nilizadeh. 2023. Demystifying NFT Promotion and Phishing Scams. *arXiv preprint arXiv:2301.09806* (2023).

[100] Doyen Sahoo, Chenghao Liu, and Steven CH Hoi. 2017. Malicious URL detection using machine learning: A survey. *arXiv preprint arXiv:1701.07179* (2017).

[101] Yun Wan, Feng Xiao, and Dapeng Zhang. 2023. Early-stage phishing detection on the Ethereum transaction network. *Soft Computing* 27, 7 (2023), 3707–3719.

[102] Colin Whittaker, Brian Ryner, and Marria Nazif. 2010. Large-scale automatic classification of phishing pages. (2010).

[103] Jiajing Wu, Qi Yuan, Dan Lin, Wei You, Weili Chen, Chuan Chen, and Zibin Zheng. 2020. Who are the phishers? phishing scam detection on ethereum via network embedding. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 2 (2020), 1156–1166.

[104] Pengcheng Xia, Haoyu Wang, Bingyu Gao, Weihang Su, Zhou Yu, Xiapu Luo, Chao Zhang, Xusheng Xiao, and Guoai Xu. 2021. Trade or trick? detecting and characterizing scam tokens on uniswap decentralized exchange. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 5, 3 (2021), 1–26.

[105] Qi Yuan, Baoying Huang, Jie Zhang, Jiajing Wu, Haonan Zhang, and Xi Zhang. 2020. Detecting phishing scams on ethereum based on transaction records. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS).* IEEE, 1–5.

[106] Zihao Yuan, Qi Yuan, and Jiajing Wu. 2020. Phishing detection on ethereum via learning representation of transaction subgraphs. In *Blockchain and Trustworthy Systems: Second International Conference, BlockSys 2020, Dali, China, August 6–7, 2020, Revised Selected Papers 2.* Springer, 178–191.

[107] Penghui Zhang, Adam Oest, Haehyun Cho, Zhibo Sun, RC Johnson, Brad Wardman, Shaown Sarker, Alexandros Kapravelos, Tiffany Bao, Ruoyu Wang, et al. 2021. Crawlphish: Large-scale analysis of client-side cloaking techniques in phishing. In *2021 IEEE Symposium on Security and Privacy (SP).* IEEE, 1109–1124.

[108] Penghui Zhang, Adam Oest, Haehyun Cho, Zhibo Sun, RC Johnson, Brad Wardman, Shaown Sarker, Alexandros Kapravelos, Tiffany Bao, Ruoyu Wang, et al. 2021. Crawlphish: Large-scale analysis of client-side cloaking techniques in phishing. In *2021 IEEE Symposium on Security and Privacy (SP).* IEEE, 1109–1124.

[109] Penghui Zhang, Zhibo Sun, Sukwha Kyung, Hans Walter Behrens, Zion Leonahenahe Basque, Haehyun Cho, Adam Oest, Ruoyu Wang, Tiffany Bao, Yan Shoshitaishvili, et al. 2022. I'm SPARTACUS, No, I'm SPARTACUS: Proactively Protecting Users from Phishing by Intentionally Triggering Cloaking Behavior. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security.* 3165–3179.