The Poorest Man in Babylon: A Longitudinal Study of Cryptocurrency Investment Scams

Anonymous Author(s)

Abstract

Governments and regulatory bodies have recognized investment scams as the most prevalent forms of cryptocurrency fraud. These scams typically use professional-looking websites to lure unsuspecting victims with promises of unrealistically high returns. In this paper, we introduce Crimson, a distributed system designed to continuously detect cryptocurrency investment scam websites as they are created in the wild. Over the first 8 months of 2024, Crimson processed approximately 6 billion domain names and classified 43, 572 unique cryptocurrency investment scam websites in real-time. Beyond detection, we provide insights into the design and infrastructure of these websites that can help users recognize scam patterns and assist hosting providers in detecting and blocking such sites. Among others, we discovered that most investment scam websites use similar templates and that 52% of all scam websites were hosted on just 10% of all resolved IP addresses, indicating a concentration of scam operations within a small subset of hosting providers. Furthermore, we investigate the inclusion of our detected scam websites in blacklists used by popular web browsers and applications, finding that the vast majority of these websites were absent. On the financial side, by analyzing the incoming transactions to scammer wallets on 6.7% of the sites detected by Crimson, we observe an estimated lower bound of 2.04M USD in losses because of cryptocurrency investment scams, pointing to tens of millions of dollars of losses in total.

ACM Reference Format:

Anonymous Author(s). 2024. The Poorest Man in Babylon: A Longitudinal Study of Cryptocurrency Investment Scams. In . ACM, New York, NY, USA, 12 pages.

1 Introduction

In September 2024, the U.S. Federal Bureau of Investigation (FBI) released its *Cryptocurrency Fraud Report 2023*, reporting on 69,000 complaints from the public regarding cryptocurrency-related financial fraud [1]. Among the various types of fraud, *investment fraud* emerged as the most common complaint, also accounting for the highest portion of reported losses. Similar statistics were reported by the Australian Competition and Consumer Commission (ACCC) [2] and U.K.'s Financial Conduct Authority (FCA) [3]. Typical cryptocurrency investment scams are propagated through professional-looking websites that promise unrealistic returns on small investments. Through sophisticated social-engineering tactics, scammers exploit the victims' trust, leading them to believe that their investments are secure, only for the victims to lose all their deposited funds.

To attract prospective victims to their sites, scammers commonly abuse popular social media platforms, usually by creating fake influencer profiles or by hacking legitimate accounts and luring the compromised account's followers to invest in cryptocurrencies [4, 5, 6]. At the same time, identifying these scammers at large remains a challenge since they can effectively hide among the hundreds of millions of legitimate users of these social media platforms. For example, Li *et al.* [7] reported that scammers oftentimes leave comments on popular YouTube channels, persuading users to invest in cryptocurrency through their scam websites. However, their analysis was limited to just 20 popular channels, and they still needed to *manually* interact with scammers before the scammers would share the URL of their cryptocurrency investment scam websites. As such, while other forms of cryptocurrency scams have been studied through prior large-scale studies [8, 9, 10], the true scale of cryptocurrency investment scam websites remains unknown.

In this paper, we introduce Crimson[1], a system that enables real-time detection of cryptocurrency investment scam websites in the wild without relying on social media and without the need to interact with scammers before they share their URLs. Crimson processes each website that is issued a TLS certificate, leveraging Certificate Transparency logs, and gradually narrows down its search to identify cryptocurrency investment scam websites through a series of filters. Eventually, each narrowed down website is validated through Meta's `Llama3:70b` and OpenAI's `GPT-4` large-language models (LLMs) using a carefully crafted prompt. This ensures that Crimson can run in a fully automated fashion without the need for any human intervention for classification. We find that the `GPT-4` model was able to correctly classify cryptocurrency investment scam websites at a 90% accuracy.

Unlike giveaway scams [10], investment scam websites are designed to mimic a credible service, typically requiring users to create an account before revealing important details such as the cryptocurrency wallet address where victims are instructed to send funds. Thus, Crimson crawls multiple pages within the investment scam websites to search for wallet addresses and other relevant information, such as email addresses and phone numbers provided by the scammers.

We utilize Crimson to conduct the first longitudinal analysis of cryptocurrency investment scams in the wild, processing billions of domain names over an 8-month period (January–September 2024). During that time, Crimson recorded 43,572 unique scam websites. We find that all detected scam websites in our dataset are hosted on 19,110 unique IP addresses, with 10% of them responsible for hosting more than half of the scam websites. This suggests that a

WWW'25, April 28 - May 02, 2025, Sydney, Australia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

[1] Crimson—CRyptocurrency InvestMent Scam detectiON

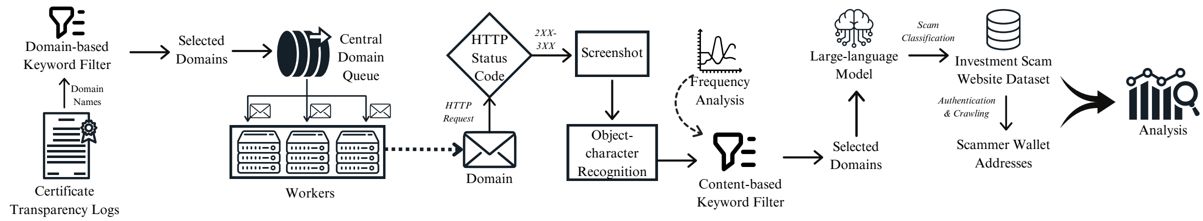


Figure 1: A schematic representation of Crimson's pipeline

relatively small number of hosts, and by extension, a limited number of hosting providers and scammers, are responsible for hosting a large portion of the identified scams. Apart from IP addresses, we cluster scam websites on the basis of their web design, JavaScript inclusions, and other information such as emails and phone numbers. Using these clustering techniques, we were able to detect commonalities between 88% of all detected investment scam websites, with a majority of scam websites belonging to more than one cluster at a time. Almost half of all detected sites remained active at the end of our observation period despite reports of fraudulent activities on social media platforms, suggesting that investment scams are persistent and, because of their facade of legitimacy, remain online over extended periods of time.

Using the extracted wallet addresses, we conduct a transaction analysis to estimate the financial losses associated with the detected scam websites. We find that 2.04M USD were sent to scammer-owned wallet addresses through Bitcoin and Ethereum payments, originating from only 6.7% of all detected scam websites. Extrapolating this number points to tens of millions of US dollars in losses across all these websites. Lastly, we report that popular blacklists are unable to provide extensive coverage of the detected cryptocurrency investment scam websites. A video demonstration of cryptocurrency investment scam websites detected by Crimson can be found here [11]. Our contributions are summarized as follows:

- We develop Crimson, the first system to detect cryptocurrency investment scam sites as soon as they are created.
- We perform detailed analysis on the detected scam websites and the financial losses that resulted from them, offering insights for cryptocurrency users and web hosting providers.
- To encourage future research in this area, we will make our dataset of detected cryptocurrency investment scam websites and related metadata publicly available upon publication of this paper. Furthermore, we will open-source Crimson's source code to enable researchers and developers to build upon our work.

2 System Design

The architecture of Crimson is illustrated in Figure 1. It comprises of six modules: ① domain selection, ② task distribution, ③ domain processing, ④ LLM-based classification, ⑤ authentication and crawling, and finally ⑥ analysis. Below, we explain each of these components in detail.

2.1 Domain Selection and Distribution

2.1.1 Certificate Transparency. Our goal is to curate a dataset of cryptocurrency investment scam websites that is representative of the true scale of the problem. As such, we need a comprehensive and reliable source of domain names. Certificate Transparency logs (CT logs) have been widely used in security research for a variety of applications involving domain names, including the detection of malicious bot activities [12], identification of cryptocurrency giveaway scams [10], and enhancement of phishing website detection [13, 14, 15]. CT is a framework designed to monitor and log the issuance of TLS certificates in a public, append-only log, facilitating the detection of unauthorized certificates. Modern web browsers enforce CT requirements by treating certificates that do not comply with these policies as untrusted, resulting in blocked connections and security warnings in the browser [16]. Given these browser policies and scammers' incentive to leverage user trust to solicit funds, it is reasonable to assume that investment scam websites will seek the issuance of TLS certificates for their domain names and that, consequently, such domains will appear in CT logs. Moreover, CT logs serve as a substantial source of domain name data, with approximately one million TLS certificates being issued and subsequently logged every hour. Therefore, utilizing CT logs can provide us with a comprehensive set of domains, ensuring that our results provide a broad view of the threat landscape of cryptocurrency investment scams. We deploy a local server using Certstream [17] to receive CT logs in real-time, capturing domain names from certificates as soon as they are issued.

2.1.2 Domain Selection. Each domain fetched from CT logs enters the first processing stage in the Crimson pipeline: Domain Selection. Initially, domains are dissected into individual keywords using a customized model based on Word Ninja [18]. Customizing the keyword segmentation model is crucial for identifying non-standard keywords, particularly those commonly found in the cryptocurrency domain, such as "eth" and "btc." For instance, `btcethinvestments[.]com` is segmented into `['btc','eth','investments']`. We compile a dataset of known cryptocurrency investment scam websites from URLscan [19] and apply our model to split the domain names into distinct keywords. We select 36 distinct keywords from the resulting keyword-set that frequently occurred in the scam dataset from URLScan. This selection process is intentionally liberal to minimize the risk of omitting potential scam domains at this phase while filtering out irrelevant websites to optimize resource usage in downstream modules. The list of keywords is provided in Appendix D. All domains having at least one keyword from the keyword-set are sent to the next processing stage. Note that we apply stemming [20] prior to comparing word lists. Stemming is a natural language process that reduces words to their base or root form. For example, the words "investors" and "investing" are both reduced to "invest", ensuring consistency in word comparison.

2.1.3 Distribution. Given that approximately 200,000 domain names pass through our *Domain Selection* stage per day, we created a distributed setup to process these domains. Selected domains are dispatched into a central queue, from which 24 worker nodes (distributed across three servers) fetch and process them. Each worker node retrieves one domain at a time from the queue and signals completion after completing all processing tasks. In case of any failure during processing, the domain is re-queued to ensure no domain is left unprocessed.

2.2 Content-based Selection

When a domain is picked up by one of the 24 workers, it undergoes a three-stage evaluation to determine its active status and relevance to cryptocurrency investments: ① *Responsiveness:* A domain entering the Crimson pipeline indicates that it has been issued a TLS certificate, but it does not guarantee that a website is operational on that domain. To account for potential delays in website responsiveness, we implement a 12-hour buffer between the pushing to the queue and when a domain becomes available to our worker nodes. To verify responsiveness after the 12-hour delay, the worker sends an HTTP request to that domain. Domains are passed onto the next check if the HTTP response has a status code within the 2XX-3XX range and are discarded otherwise. ② *Screenshot:* A full-page screenshot of the website is captured using Selenium [21] and used for analysis. ③ *Text extraction:* Using an Object Character Recognition (OCR) tool [22], we extract the text from the captured screenshot. This method is advantageous over merely retrieving HTML, as it also captures text from images and JavaScript-rendered elements, which play a role in the professional theme of the website used to gain the victims' trust.

The text retrieved using OCR is tokenized, *stemmed* into individual keywords, and compared against a second keyword list composed of words commonly found in the homepage text of known cryptocurrency investment scam websites from URLScan. These keywords are grouped into three categories: *Investment*, which includes variations of the word "invest"; *Coins*, which includes cryptocurrency-related terms like "btc" and "btcusd"; and *Context*, which consists of commonly used scam-related terms such as "deposit", "withdraw", as well as words designed to create a sense of urgency, trust, and legitimacy, such as "secure" and "safe." If the stemmed OCR-generated tokens contain a word from each of these groups, it is forwarded to the next processing stage. As Crimson identifies scam websites over time, we perform frequency analysis to find any recurring words missing from our keyword list, adding them to refine the keyword list. The final list includes a total of 82 keywords and is provided in Appendix D.

2.3 LLM-assisted Classification

Previous studies have primarily relied on human intervention to categorize malicious web pages [23, 10]. However, manual categorization is impractical for Crimson, given the total of approximately 320,000 domains processed through the content-based selection module. In contrast, recent advances in large-language models (LLMs) have demonstrated their effectiveness in automating repetitive tasks like code analysis, penetration testing, and phishing detection [24, 25, 26, 27, 28].

Table 1: Performance and cost efficiency of LLMs in classifying cryptocurrency investment scam websites. The hybrid approach using `GPT-4 + LLama3:70b` offers a balance between performance and cost.

LLM	Accuracy (%)	Estimated Total Cost (USD)
GPT-4	90	1200
GPT-4 + Llama3:70b	**88**	**130**
Llama3:70b	87	0
GPT-4 Vision	79	1800
Mistral:7b	67	0
Llama2:70b	59	0
Llama2:13b	54	0

Therefore, as a final step towards scam website detection, we utilize LLMs for classifying websites as *scam* or *not scam*. To choose the appropriate LLM for our task, we choose from a pool of 6 popular LLMs available at the time, which are listed in Table 1.

To evaluate the performance of each LLM, we manually categorized a random sample of 300 websites that had passed the content-based filter, with half classified as scams and the other half as non-scams. Each LLM was tasked with classifying these websites using a uniform prompt, provided in Appendix B. Along with the prompt, we provided OCR-extracted text for the LLMs to reference when making their classifications. For the `GPT-4 Vision` model, we supplied a full-page screenshot instead of text, as it can process visual content directly. The accuracy and total estimated cost of each of the LLMs evaluated are presented in Table 1. Our findings show that the `GPT-4` model yielded the most accurate classifications. However, in terms of operational cost, the `GPT-4` and `GPT-4 Vision` models each cost 10 USD per 1M input tokens and 30 USD per 1M output tokens. Meanwhile, Meta's `Llama3:70b` model offers a cost-effective solution and achieves 87% accuracy—comparable to `GPT-4`—but occasionally fails to provide a conclusive *yes* or *no* answer, a limitation not observed in `GPT-4`. To balance accuracy with cost-effectiveness, we adopt a hybrid approach and primarily use the `Llama3:70b` model for classification. In cases where it produced inconclusive results, we route the query to the `GPT-4` model for final verification, yielding an overall accuracy of 88%.

2.4 Account Creation and Wallet Extraction

Once scam websites are identified using LLMs, Crimson automates the process of retrieving scammer-owned cryptocurrency wallet addresses where victims are instructed to send funds. This requires interacting with the websites as a typical user, which includes signing up, logging in, and navigating the site's internal pages. After logging in, users are usually presented with investment plans, cryptocurrency types, and wallet addresses for transferring funds. This requirement for authentication is unique to investment scams, as scammers aim to make their services appear more legitimate. In contrast, other forms of cryptocurrency scams, such as giveaway scams [10], typically display the wallet address directly on the homepage, making it easier to crawl scammer wallet addresses.

Crimson first navigates to the inner pages of the website to locate a sign-up page. Pages containing two or more password fields or specific keywords in the URL are identified as potential sign-up pages. Once a sign-up page is detected, Crimson fills in the form fields, typically including first name, last name, country of residence, username, password, and other non-text elements like

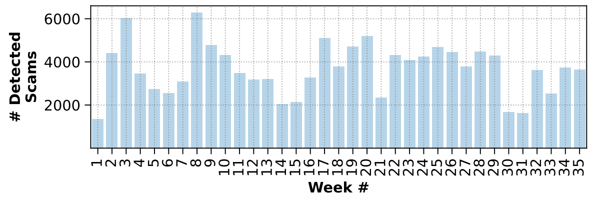


Figure 2: Weekly number of unique scam websites discovered by Crimson

dropdowns or radio buttons. To generate unique email addresses for each registration, we use Mailinator [29], leveraging a dictionary of common first and last names combined with a private email domain. This approach ensures that email addresses are not overly reused, reducing the likelihood of raising suspicion among colluding scammers. If scammers detect repeated use of the same email addresses, they could block our connections to their websites, which would risk losing access to future scam websites they might create and the associated wallet addresses. After completing the form, Crimson submits the information to complete the sign-up process.

Next, Crimson identifies the login page and logs in using the credentials created during the sign-up process. Once logged in, scam websites typically redirect users to a dashboard where they can view various investment packages, withdraw or transfer funds, and, most importantly, deposit funds. Crimson directs to the deposit page of the scam website and uses `iocsearcher` [30] tool to extract wallet addresses. It also takes Screenshots and stores the HTML code of each page visited during crawling.

Automating the authentication process in arbitrary websites is a known difficult problem [31]. For instance, many require solving CAPTCHAs, or contain a secondary authentication process where users will have to wait till the website administrators provide them access to the user's profile and dashboard. When automated crawling successfully authenticates to a website but fails to retrieve the wallet address, we manually navigate through the site's structure by utilizing credentials Crimson used to login to the same website to extract this critical information. Even when Crimson successfully logs into scam websites, finding the wallet address can be difficult as it is often deep within the site, requiring navigation through multiple links or dropdown menus. We also encountered instances where the deposit page did not reveal the associated wallet address due to website bugs. These complexities result in a low success rate in automatically identifying wallet addresses from scam websites. However, the wallet addresses that we do collect still reveal substantial financial losses tied to investment scams (§4).

3 Scam Website Analysis

In this section, we provide a comprehensive analysis of our dataset of investment scam websites and associated wallet addresses.

From the approximately 25 million domain names that Crimson parses daily through CT logs, it detects an average of 189 unique cryptocurrency investment scam websites each day. Figure 2 shows the weekly number of unique websites that Crimson detected over our 8-month deployment period. We identify a total 43,572 unique cryptocurrency investment scam domain names (comprising of 38,365 unique second-level domains) over the first 8 months of 2024, resolving to 19,110 unique IP addresses.

Table 2: Shared IP addresses hosting multiple scam websites, along with the largest cluster sizes. Certain hosting providers are hosting large numbers of scam websites, often on the same IP address.

Hosting Provider	Total Shared IPs (T=4,900)	Largest Cluster	Total IPs (T=19,110)	Total Websites (T=43,572)	Perc. in Top 10k
Hostinger	2,149 (44%)	55	5,597 (29%)	11,160 (26%)	1%
Cloudflare	694 (14%)	327	7,207 (38%)	9,296 (21%)	2%
NameCheap	209 (4%)	23	729 (4%)	1,094 (3%)	6%
OVH	170 (3%)	214	312 (2%)	2,826 (6%)	1%
Hetzner	120 (2%)	199	338 (2%)	1,998 (5%)	4%
Interserver	117 (2%)	148	165 (1%)	1,084 (2%)	0%
Amazon	71 (1%)	197	340 (2%)	1,384 (3%)	4%
Contabo GmbH	68 (1%)	55	159 (1%)	607 (1%)	0%
20I Limited	55 (1%)	29	71 (1%)	528 (1%)	0%
WHG	51 (1%)	18	90 (1%)	296 (0%)	0%
(Total)	*3,704 (76%)*		*15,008 (79%)*	*30,273 (69%)*	*~16%*

3.1 Domain Name Characteristics

We analyze the domain name characteristics of cryptocurrency investment scam websites, focusing on both top-level and second-level domains. For comparative purposes, we collect data on the top-level domains (TLDs) used by the top 10K websites listed by Tranco [32] as of August 25, 2024.

In terms of the most popular TLDs, we find that a majority of websites in both datasets use common TLDs like *.com*, *.net*, and *.org*, with *.com* being the most prevalent, accounting for 64% of all scam websites and 50% of the top 10K dataset. The widespread use of these reputable TLDs by scammers suggests that they are willing to pay a premium to create a sense of legitimacy and increase trustworthiness with potential victims. Beyond top-level domains, we examine second-level domain name patterns and observe that only 10 words from the *Domain Selection* module's keyword list account for 71% of scam websites, compared to 0.6% of websites in the Tranco top 10K. Tables 6 and 7 in Appendix C.1 list the number of websites across the most popular top- and second-level domains.

3.2 Clustering

3.2.1 IP-based Clustering. Crimson resolved the detected 43,572 investment scam websites and identified 19,110 unique IP addresses. Our analysis reveals that of all detected investment scam websites, 29,300 (67%) of them share only 4,900 (26% of all) IP addresses. Figure 3 shows the cumulative percentage of all websites that were hosted over the cummulative percentage of all IP addresses. We can therefore conclude that more than half of the scam websites are associated with merely 10% of the IP addresses in our dataset. This disproportionate concentration of scam websites on a small subset of IP addresses suggests the use of shared hosting infrastructure among scammers as well as takedown opportunities for defenders.

Furthermore, we observe that a significant portion of websites sharing IP addresses are concentrated among a few hosting providers. Specifically, 10 hosting providers account for 76% of all shared IP addresses. Table 2 provides a breakdown of these shared IP addresses by hosting provider, including the size of the largest cluster, total number of websites hosted by each provider, and the total number of unique IP addresses they resolved to. Hostinger was responsible for sharing 2,149 IP addresses among different scam websites, the highest among all hosting providers.

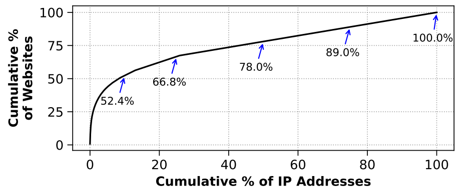


Figure 3: Cumulative Distribution of investment scam websites over IP Addresses. A majority of scam websites are hosted on a small percentage of all IP addreses

The most concentrated cluster was 55 scam websites hosted on a single IP address. Overall, Hostinger accounted for approximately one-quarter of all detected scam websites. Notably, 8 out of these 10 hosting providers in Table 2 have a median of 3 or more websites sharing the same IP address.

Even though the hosting provider *Unified Layer* was responsible for hosting only 390 scam websites over 32 IP addresses, it hosts the largest cluster of websites sharing a single IP address, with 346 websites resolving to the same IP, among which 301 belong to the same second-level domain. The next largest cluster is hosted by Cloudflare, with 327 websites sharing an IP address. Interestingly, all these websites are subdomains of `cryptocurrencies-offers.com`. Each of these subdomains hosted a cryptocurrency investment scam, using multiple website templates to propagate the same fraudulent scheme, attempting to convince potential victims into trading cryptocurrencies on their platform with promises of high returns.

In total, 43,572 investment scam websites were hosted by only 874 web hosting providers. Similar to domain names, there is a notable difference in the web hosting providers chosen for scam websites compared to those used by Tranco's top 10K websites. Providers such as Hostinger, OVH, and Hetzner host a significant portion of scam websites, likely due to their low-cost hosting services. These affordable solutions appear to be attractive to scam operators who seek to minimize hosting expenses while maximizing their profits from scamming.

3.2.2 Web Design-based Clustering. To systematically evaluate the design of the investment scam websites, we employ the perceptual hash (p-hash) algorithm to generate a 64-bit fingerprint for each website screenshot. Identical p-hash values indicate exact visual similarity between screenshots. However, to cluster screenshots of websites with similar designs but not identical, we group images whose p-hashes differ by a Hamming distance of up to 8 bits–that is, images sharing at least 56 our of 64 bits in their p-hash. We determined this threshold through iterative experimentation and manual analysis, allowing the clustering algorithm to group screenshots that share the same overarching template while permitting minor variations such as background color, investment plans, titles, and (fake) customer reviews.

Our analysis reveals that investment scammers often design their websites to appear professional and trustworthy, aiming to convince potential investors of their legitimacy. Common features of the investment scam sites detected by Crimson include attractive and modern interfaces, detailed contact information, live chat services, fabricated user reviews, false notifications of high-profit withdrawals by other investors, misleading statistics indicating substantial earnings, and counterfeit business certificates. After introducing their platforms, scammers typically present users with various investment plans that specify minimum and maximum investment amounts, promise unrealistically high returns on investment, and outline short time-frames for these returns to materialize. An example of a counterfeit certificate displayed on the investment scam website `alpinextrade.com`, detected by Crimson, is available in Figure 6 (Appendix C.2).

Table 3: Distribution of shared IoCs among detected cryptocurrency investment scam websites. Scammers are reusing the email addresses, phone numbers, and social media handles across websites.

Identifier	# Websites	Reused	Largest Cluster
Email Address	27,036 (58%)	1,806	318
Phone Number	12,092 (27%)	1,411	121
Telegram Handle	4,293 (10%)	392	189
Twitter Handle	4,014 (9%)	367	126
Facebook Handle	3,467 (8%)	284	389
Instagram Handle	2,999 (7%)	305	185
GitHub Handle	1,635 (4%)	90	719
LinkedIn Handle	1,409 (3%)	157	50
Ethereum Address	1,334 (3%)	78	26
Bitcoin Address	1,278 (2%)	34	20
YouTube Channel	939 (2%)	122	24
Tronix Address	408 (1%)	13	19

Applying our aforementioned clustering methodology, we grouped a total of 17,285 investment scam website *home-page* screenshots–representing 40% of all detected scam websites–into 4,335 clusters, each containing at least two web-pages. The largest cluster comprises 347 scam website screenshots, and each of the top five clusters contain more than 120 screenshots. To validate the accuracy of our clustering algorithm, we randomly sampled 100 clusters of varying sizes and manually inspected the screenshots. Our inspection confirmed that all screenshots within each cluster shared the same website template with only minor alterations. A few examples of investment scam web-designs are provided in Figure 7 (Appendix C.2).

Additionally, we applied the same clustering approach to inner page screenshots, such as the "About Us", account settings, login/sign-up, investment plan dashboard, and investment deposit pages. This analysis resulted in 14,042 inner-page screenshots from 2,172 domains (4% of all detected scams) being grouped into 3,732 clusters. Interestingly, we identified 682 cases where websites that were not clustered together based on their home page designs were found in the same clusters when analyzing inner pages. This suggests that scammers may reuse templates for specific sections of their websites, such as post-authentication dashboards and login pages, even if their home pages differ in design.

3.2.3 IoC-based Clustering. To establish credibility, scammers make use of tactics to portray legitimacy and trustworthiness to potential victims. To systematically identify and analyze these deceptive practices, we employed `iocsearcher` by Caballero *et al.* [33, 30] to extract various indicators of compromise (IOCs) directly from the HTML code of all scam websites from their home- and inner-pages.

Table 3 lists the specific indicators extracted across all detected scam websites, along with the count of the websites from which each indicator was retrieved. Moreover, it lists the total number of identifiers that are shared among more than one website from

Table 4: Most common JavaScript library use-cases

Use Case	# Scam Websites	Perc. in Top 10K
JQuery	35,597 (82%)	36%
Language Translators	18,828 (43%)	5%
Trading View	15,260 (35%)	1%
Fake Notification	3,296 (8%)	1%
Chat Services	1,279 (3%)	2%

each category and the largest cluster of websites that share the same identifier. Note that since the identifiers are extracted from the raw HTML code, it is not necessary that it will appear on the website view. Rather, as oftentimes scammers lazily reuse templates to distribute their scam over a large number of domains, these identifiers can also be present inside the commented-out portions of their code.

The top three most shared phone numbers among websites were shared between 121, 78, and 74 websites. Interestingly, the phone number `+19043263***` was shared among 74 websites, and upon further investigation, we found a related report on `scam-detector.com` where, in October 2023, a user had reported being scammed by an investment website providing this exact number [34]. According to the report, the victim was deceived into sending the equivalent of 1,346 USD through Binance to an investment scam website under the guise of promised profits. The scam domain reported by the user is no longer active. However, multiple domains found by Crimson that have the same phone number are still active at the time of writing. Moreover, the email address `support@brynamics.xyz` was found to be shared among 184 scam websites. A basic online search of this email yields numerous results linked to various investment scam websites that have advertised it. We provide further examples of common YouTube channels, Instagram accounts, and Telegram handles that we found in Appendix C.3.

Upon inspection, we found that the shared Ethereum, Bitcoin, and Tronix wallet addresses detected on scam website homepages were often dummy addresses, typically displayed alongside fabricated metrics such as large cryptocurrency withdrawals by supposed users. This was done to create the illusion that the scammers' service was trustworthy and actively used. However, the actual wallet addresses, where victims were instructed to send funds, were generally revealed only after the victims had logged in.

3.2.4 JavaScript-based Clustering. Apart from replicating common website designs, scammers often reuse common JavaScript elements across websites. When Crimson detects a scam site, it extracts all JavaScript inclusions, whether local or remote, from the HTML code. The most frequently encountered JavaScript inclusions are listed in Table 4. While these JavaScript libraries may not be inherently malicious, understanding their usage patterns helps us identify the common elements of investment scam websites and how scammers exploit them to increase their reach, lend their sites an appearance of legitimacy, and manipulate users into sending them funds.

Our analysis reveals that 81% of scam websites are using JQuery. Its features provide scammers with capabilities for manipulating the Document Object Model (DOM) and thus enhancing website interactivity. For instance, with a few lines of JQuery, scammers have set up chatbots to facilitate basic functionalities such as text-based interactions and predefined responses. Additionally, we observe third-party JavaScript inclusions for chat services on 3% of the detected scam websites. Beyond JQuery, language translators and trading view widgets are the next most common JavaScript inclusions observed on scam websites. Translator scripts enable scammers to automatically localize their content based on user-selected languages or even adapt dynamically to the user's browser or device language settings. This tactic creates a more personalized experience that mimics the behavior of legitimate websites and also facilitates scammers to reach a broader audience across different countries and languages, increasing the pool of potential victims. Trading view widgets embed real-time market data, charts, and financial information into websites. Scammers exploit these widgets to create trading dashboards, price tickers, and candlestick charts. Moreover, scammers employ the social engineering tactic of fake notifications to further manipulate user behavior. These notifications often mimic real-time updates about other users' activities—such as deposits, withdrawals, or profits earned—aiming to create a sense of social proof. For example, on March 28 2024, Crimson detected the scam website `securedcryptoassets.com`, which periodically displayed a notification stating: *"Dustin from Anaheim just earned $41,851 25 minutes ago."*. This notification would refresh every few seconds with a random name, city, and earnings.

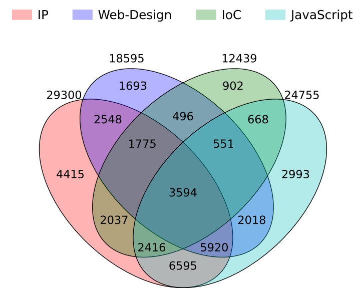


Figure 4: 4D Venn diagram illustrating the numbers of scam websites shared amongst all cluster permutations

3.2.5 Cluster Overlap Analysis. If all websites that belong to at least one cluster are put together, it accounts for 88% of all detected scam sites. Figure 4 shows a four-dimensional overlap between the websites in each cluster, showing that a majority of websites are present in more than one cluster at once. These findings do not include the websites that belonged to the JQuery cluster since JQuery is a general framework that was found in 36% of the top 10K Tranco sites.

3.3 Life-span

During our data-collection period, we monitored each identified scam domain on a daily basis to determine whether it continued to host an investment-scam website. While an unresponsive website guarantees that the scam content is no longer available, the inverse is not true. That is, the web-server associated with a scam website can still be returning content in the days and weeks following its first discovery without that content being necessarily associated with the initial scam (*e.g.,* serving entirely different content, takedown notices from the hosting providers, etc.).

To address this ambiguity, we utilize the title of the website to determine whether it continues to host the same content. If the

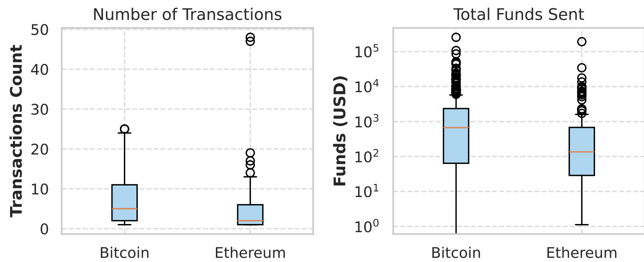


Figure 5: Total number of transactions and US Dollars directed towards scammer-owned cryptocurrency wallets from each victim

cosine similarity [35] between the current and initial title exceeds 0.9, we consider the website to still be hosting scam-related content.

Crimson identifies a daily average of 522 scam websites, of which 189 are unique. The total number of scams detected each day also includes websites that were previously identified; this repetition occurs because these domains reappear in CT logs, leading to their re-detection by Crimson. These duplicate CT logs can occur due to multiple scenarios, such as obtaining a new certificate after a domain has changed hosting providers following a take-down by the previous hosting provider, upcoming certificate expirations, or when a new certificate is issued for an additional server that hosts the same website content. We observe that 29,340 scam domains appeared more than once in CT logs, thus resulting in a reappearance in our dataset. Our analysis reveals substantial variability in the operational status of cryptocurrency investment scam websites throughout the 8 months. Specifically, 17,653 websites exhibited intermittent periods of inactivity followed by reactivation after short periods of time. Interestingly, 2,355 websites reactivated with a different hosting provider subsequent to a period of inactivity. This pattern likely suggests that these sites were shut down by their previous hosts, prompting the scammers to obtain web hosting from new providers in order to make their scam sites available again. We find that 16,635 (44%) of websites remained consistently inactive or underwent content modifications during the final ten days of monitoring. In contrast, 22,983 (47%) websites continued to be active and were still hosting cryptocurrency investment scams by the end of the observation period.

4 Estimating Financial Losses

In this section, we describe our methodology and results for estimating the success of scammers in terms of receiving cryptocurrency transactions in the wallet addresses they provided on their scam websites. Given the predominant market capitalization and the vast user bases of Bitcoin and Ethereum, we choose to focus our analysis on these two cryptocurrencies.

We collected 489 Ethereum addresses and 1,106 Bitcoin wallet addresses from 2,923 scam websites. We then aggregated all incoming transactions to all wallet addresses by leveraging blockchain explorers [36, 37]. For Bitcoin, we aggregate the total amount directed to scammers' wallet addresses by summing the Bitcoins from the output slot of each transaction where the address belongs to a scammer. This approach ensures we avoid over-counting funds in cases where change addresses return a portion of the funds back into the victim's wallet. For Ethereum, we record the aggregate amount sent where the scammer's wallet is the receiving address.

To further mitigate the risk of overestimation in our results, we adhere to the guidelines provided by Gomez *et al.* [38] and convert the resulting amounts of each cryptocurrency to US dollars based on the adjusted closing prices on the days the transactions occurred to provide a tight estimation of the monetary value victims transferred using each transaction. Furthermore, we exclude any transactions that occurred before the creation date of the scam websites (per WHOIS data), displaying the wallet addresses to ensure that our revenue estimations only reflect transactions influenced by the scam websites themselves. That is, if a specific cryptocurrency scam website discovered by Crimson has a domain registration date of June 1, 2024, we exclude cryptocurrency transactions predating that registration date.

Our results, shown in Figure 5, clearly reveal that scammers have unfortunately been successful in their goal to lure victims into sending funds into their wallets. We find that 3,497 transactions were sent towards scammer-owned wallets by 189 unique Ethereum wallet addresses and 1,907 Bitcoin senders. In total, transactions sent towards scam wallets sum up to 2.04M US dollars, with 83% of payments sent through Bitcoin. Even though the captured financial losses represent a significant amount of funds, they stem from only 6.7% of the total scam websites that we identify. This is primarily due to the challenges we face in successfully authenticating and extracting wallet addresses from many sites. We suspect that the actual financial losses linked to these scams are much higher than reported. On average, approximately 3.6K USD were sent to scammer wallet addresses. If this average is extrapolated to all of the detected scam websites, we estimate the total financial loss to be more than 100M USD. Tables 12 and 13 in Appendix E list the top earning investment scam domains through both Bitcoin and Ethereum payments, along with their estimated revenues and operational status.

We source a list of 795 custodial wallet addresses belonging to online exchanges such as Coinbase [39] and Binance [40], from Etherscan [36] and CoinCarp [41] and find that only 13% of all transactions to scammers were sent through them. This indicates that non-custodial wallets are more popular among victims of investment scams and that such financial losses could be avoided if non-custodial wallets detect scam websites and warn users in a timely manner. One of the most popular non-custodial wallet providers, MetaMask [42] makes use of blacklists, displaying a warning when users visit a known malicious website. We discuss the coverage of this blocklist, along with warning services by other providers in the following section.

Lastly, we briefly report on the phenomenon of multiple *different* users being shown the same wallet address for depositing funds on the same investment site. Having all users deposit funds to the same address would make it difficult for a legitimate service to determine which funds came from which user and would allow malicious users to frontrun other users' funds. As such, we argue that a shared wallet address between unrelated users on the same site is one more indicator that the site is a scam with no intention of ever correctly tracking user funds. To this end, we used Crimson *twice* on 15 randomly sampled websites where our tool was initially able to create accounts and identify wallet addresses. For all 15 sites, we observed the same wallet address being shown to our two different Crimson-generated accounts.

Table 5: Percentage of scam websites detected by Crimson that were also present in known blacklists.

Name	Intersection
VirusTotal [43]	20%
Metamask [42]	2%
SEAL-ISAC [44]	2%
Google Safe Browsing [45]	1%
WalletGuard [46]	1%
Phishfort [47]	1%
ChainPatrol [48]	1%

5 Discussion

5.1 Coverage

Due to the large number of scam websites in the wild, various blacklists have been developed to flag suspicious websites and warn users within their web browsers of potential malicious activity. For example, all modern browsers utilize Google Safe Browsing (GSB) to display a warning page when users are about to visit a suspicious website. The Metamask wallet extension offers similar functionality to GSB that is specific to cryptocurrency-related malicious websites. Services such as WalletGuard, Phishfort, ChainPatrol, and SEAL-ISAC also provide blacklists for malicious websites and can be integrated into web browsers and third-party applications.

To understand to what extent these existing blacklists and extensions are able to protect users from cryptocurrency investment scams, we check whether our Crimson-discovered websites are flagged as malicious. Specifically, we compare the coverage of 7 services against a random sample of one-third of the Crimson dataset, and list our results for each blacklists in Table 5. Evidently, a majority of the existing blacklists were not able to provide a reasonable detection rate of investment scam websites. We suspect that this low coverage in existing blacklists is due to the general nature of websites they aim to detect. Going forward, these services could use a Crimson-like system to augment their detection logic and improve their performance.

5.2 Limitations

While Crimson was able to find tens of thousands of real scam investment sites, it suffers from some limitations. ① Even though our keyword filters were carefully selected and optimized for identifying cryptocurrency investment scams, if attackers deliberately avoid using these targeted keywords or adapt their content to bypass the filters, their scam websites could evade detection. However, it is also hard for attackers to avoid our keywords since it would be challenging to sell cryptocurrency investment sites without using them. That is, complete avoidance of investment and cryptocurrency terms will also reduce (if not altogether eliminate) their conversion rates for victim users. If necessary, our keyword filters can always be expanded with additional terms to identify such future websites at the expense of extra resources to handle the increased workload and LLM usage. ② Our use of LLMs for scam validation was motivated by the goal of eliminating the need for human intervention in classifying scams. However, this approach comes with a trade-off in terms of reduced accuracy compared to manual methods. As LLMs continue to evolve and improve, we anticipate that their accuracy in detecting scams will increase over time, allowing Crimson to more reliably identify fraudulent websites. ③ Since each website has a different authentication template, and many employ CAPTCHA or other mechanisms to block automated access, automating the sign-up and login process to crawl wallet addresses is difficult. Despite these hurdles, we were still able to extract thousands of wallet addresses from post-authentication dashboards and estimate overall financial losses.

6 Related Work

Prior to our work, studies on cryptocurrency investment scam websites relied on links collected through social media and online forums such as YouTube and Bitcointalk [49, 50, 7]. Several research efforts have focused on Ponzi schemes, which are a broader category of cryptocurrency investment scams. These have been extensively studied on blockchain platforms like Ethereum and Bitcoin, where researchers utilize blockchain data—such as smart contracts and transaction patterns—to analyze their structure and operation [51, 52, 53, 54, 55]. For instance, Bartoletti *et al.* [52] conducted a comprehensive survey of "Smart" Ponzi schemes on Ethereum, examining how these schemes leverage smart contracts to automate their fraudulent processes. To the best of our knowledge, Crimson is the first system built to detect cryptocurrency investment scam websites in real-time immediately upon their creation.

In 2023, Li *et al.* [10] developed CryptoScamTracker, a system for identifying cryptocurrency giveaway scams [56, 57], where scammers deceive victims by promising to return a multiplied amount of cryptocurrency if they first send a small amount to a provided wallet, often posing as a donation event or falsely advertising giveaways endorsed by public figures. Comparing our results, we find that investment scams are more widespread, with approximately 8.5x more scam websites detected overall. Additionally, we find that investment scam websites remain active for longer durations and estimate that they are responsible for significantly larger financial losses, a pattern also reported by government agencies [1, 2]. In addition, the security community has examined other forms of social engineering-based cryptocurrency fraud, including pump-and-dump schemes [58, 59], technical support scams [8], NFT-related frauds [60, 61, 62], YouTube comment scams [7], and sextortion [63].

7 Conclusion

In this paper, we developed and utilized Crimson to detect 43,572 cryptocurrency investment scam websites over a period of 8 months. We conducted an in-depth analysis of these scam websites, clustering them based on their IP addresses, web design, JavaScript inclusions, and other relevant information present in the HTML. Our findings revealed that 47% of all detected scam websites were still active at the end of our data collection period and that scam websites often reappear through different hosting providers if they are taken down. In terms of financial losses, we estimated a lower bound of 2.04M USD sent to scammer-owned wallet addresses by victims. Lastly, we identified that popular blacklists used by web browsers and applications do not provide adequate coverage for investment scam websites, leaving many undetected and, as a result, failing to issue warnings to users who visit them.

Availability. We will be open-sourcing Crimson and our collected data upon publication of this paper.

References

[1] Federal Bureau of Investigation. *Federal Bureau of Investigation Cryptocurrency Fraud Report 2023.* https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3CryptocurrencyReport.pdf. October 2024.

[2] Australian Competition and Consumer Commission. *ACCC calls for united front as scammers steal over $3bn from Australians.* https://www.accc.gov.au/media-release/accc-calls-for-united-front-as-scammers-steal-over-3bn-from-australians?utm_source=twitter_accc&utm_medium=social&utm_campaign=p_tru_g_awa_c_scams&utm_content=targeting_scams_2023&sf176457037=1. October 2024.

[3] Financial Conduct Authority (FCA). *Over £27 million reported lost to crypto and forex investment scams.* https://www.fca.org.uk/news/press-releases/over-27-million-reported-lost-crypto-and-forex-investment-scams. October 2024.

[4] Federal Trade Commission (FTC). *Social media: a golden goose for scammers.* https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2023/10/social-media-golden-goose-scammers. October 2024.

[5] Norton. *12 Instagram scams to know and avoid in 2024.* https://us.norton.com/blog/online-scams/instagram-scams. October 2024.

[6] Charlelie (Ciicii) Mavusi. *My Instagram account was hacked and I was tricked into promoting scam cryptocurrency.* https://www.getsafeonline.org/personal/blog-item/my-instagram-account-was-hacked-and-i-was-tricked-into-promoting-scam-cryptocurrency/. October 2024.

[7] Xigao Li, Amir Rahmati, and Nick Nikiforakis. "Like, Comment, Get Scammed: Characterizing Comment Scams on Media Platforms". In: *Proceedings Network and Distributed System Security Symposium.* 2024.

[8] Bhupendra Acharya et al. *Conning the Crypto Conman: End-to-End Analysis of Cryptocurrency-based Technical Support Scams.* 2024. arXiv: 2401.09824 [cs.CR]. URL: https://arxiv.org/abs/2401.09824.

[9] P. Xia et al. "Characterizing Cryptocurrency Exchange Scams". In: *Computers & Security* 98 (2020), p. 101993.

[10] Xigao Li, Anurag Yepuri, and Nick Nikiforakis. "Double and nothing: Understanding and detecting cryptocurrency giveaway scams". In: *Proceedings of the Network and Distributed System Security Symposium (NDSS).* 2023.

[11] Anonymous Authors. *The Poorest Man in Babylon: Demonstration Video.* https://anonymous.4open.science/api/repo/scam-website-9EE7/file/scam-website.mp4?v=bc004592. October 2024.

[12] Brian Kondracki, Johnny So, and Nick Nikiforakis. "Uninvited guests: Analyzing the identity and behavior of certificate transparency bots". In: *31st USENIX Security Symposium (USENIX Security).* 2022, pp. 53–70.

[13] Mashael AlSabah et al. "Content-agnostic detection of phishing domains using certificate transparency and passive dns". In: *Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses.* 2022, pp. 446–459.

[14] Edona Fasllija, Hasan Ferit Enişer, and Bernd Prünster. "Phish-Hook: Detecting phishing certificates using certificate transparency logs". In: *Security and Privacy in Communication Networks: 15th EAI International Conference, SecureComm 2019, Orlando, FL, USA, October 23–25, Proceedings, Part II 15.* Springer. 2019, pp. 320–334.

[15] Brian Kondracki et al. "Catching transparent phish: Analyzing and detecting mitm phishing toolkits". In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* 2021, pp. 36–50.

[16] Google. *Certificate Transparency in Chrome.* https://googlechrome.github.io/CertificateTransparency/. July 2024.

[17] CaliDog. *certstream-server.* https://github.com/CaliDog/certstream-server. October 2024.

[18] keredson. *Word Ninja.* https://github.com/keredson/wordninja. July 2024.

[19] URLScan. *URLScan.* https://urlscan.io/. July 2024.

[20] IBM. *What is stemming?* https://www.ibm.com/topics/stemming. October 2024.

[21] Selenium. *WebDriver.* en. https://www.selenium.dev/documentation/webdriver/. September 2024.

[22] tesseract-ocr. *tesseract.* https://github.com/tesseract-ocr/tesseract. July 2024.

[23] Pieter Agten et al. "Seven Months' Worth of Mistakes: A Longitudinal Study of Typosquatting Abuse." In: *NDSS.* 2015.

[24] Peiyu Liu et al. "Exploring {ChatGPT's} Capabilities on Vulnerability Management". In: *33rd USENIX Security Symposium (USENIX Security).* 2024, pp. 811–828.

[25] Gelei Deng et al. "{PentestGPT}: Evaluating and Harnessing Large Language Models for Automated Penetration Testing". In: *33rd USENIX Security Symposium (USENIX Security).* 2024, pp. 847–864.

[26] Rukhshan Haroon and Fahad Dogar. "TwIPS: A Large Language Model Powered Texting Application to Simplify Conversational Nuances for Autistic Users". In: *arXiv preprint arXiv:2407.17760* (2024).

[27] Ruofan Liu et al. "Less Defined Knowledge and More True Alarms: Reference-based Phishing Detection without a Pre-defined Reference List". In: *33rd USENIX Security Symposium (USENIX Security).* 2024, pp. 523–540.

[28] Yuexin Li et al. "KnowPhish: Large Language Models Meet Multimodal Knowledge Graphs for Enhancing Reference-Based Phishing Detection". In: *arXiv preprint arXiv:2403.02253* (2024).

[29] Mailinator. *Mailinator.* https://www.mailinator.com/. July 2024.

[30] Malicia Lab. *iocsearcher.* https://github.com/malicialab/iocsearcher. September 2024.

[31] Kostas Drakonakis, Sotiris Ioannidis, and Jason Polakis. "The cookie hunter: Automated black-box auditing for web authentication and authorization flaws". In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* 2020, pp. 1953–1970.

[32] Victor Le Pochat et al. "Tranco: A research-oriented top sites ranking hardened against manipulation". In: *arXiv preprint arXiv:1806.01156* (2018).

[33] Juan Caballero et al. "The rise of goodfatr: A novel accuracy comparison methodology for indicator extraction tools". In: *Future Generation Computer Systems* 144 (2023), pp. 74–89.

[34] Anup Kumar Sharma. *Anup Kumar Sharma October 23, 2023 at 8:56 am.* https://www.scam-detector.com/crypto-info/. October 2024.

[35] Educative.io. *How to find similarity between two words using NLP.* https://www.educative.io/answers/how-to-find-similarity-between-two-words-using-nlp. October 2024.

[36] Etherscan. *Etherscan.io.* https://etherscan.io/. September 2024.

[37] Blockstream. *Bitcoin Explorer - Blockstream.info.* https://github.com/Blockstream/esplora/blob/master/API.md. September 2024.

[38] Gibran Gomez, Kevin van Liebergen, and Juan Caballero. "Cybercrime bitcoin revenue estimations: Quantifying the impact of methodology and coverage". In: *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security.* 2023, pp. 3183–3197.

[39] Coinbase. *coinbase.* https://www.coinbase.com/. October 2024.

[40] Binance. *binance.* https://www.binance.com/. October 2024.

[41] CoinCarp. *Bitcoin(BTC) Exchange Wallet Address List and Balance Change | CoinCarp.* https://www.coincarp.com/currencies/bitcoin/exchange-wallets/. September 2024.

[42] Metamask. *The ultimate crypto wallet for DEFI, Web3 apps, and NFTs | MetaMask.* en. URL: %5Curl%7Bmetamask.io%7D.

[43] VirusTotal. *VirusTotal API v3 Overview.* en. https://docs.virustotal.com/reference/overview. September 2024.

[44] SEAL-ISAC. *An ISAC Tailor-Made for Crypto, Blockchain, and Web3.* https://isac.securityalliance.org/. October 2024.

[45] Google. *Safe Browsing – Google Safe Browsing.* en. https://safebrowsing.google.com/. September 2024.

[46] WalletGuard. *Protect your Crypto | Wallet Guard.* September 2024. URL: %5Curl%7Bhttps://www.walletguard.app/%7D.

[47] Phishfort. *Protect your brand and revenue from attacks.* https://www.phishfort.com/. October 2024.

[48] Chainpatrol. *Real-time Brand Protection for Leading Web3 Companies.* https://chainpatrol.io/. October 2024.

[49] Gilberto Atondo Siu et al. ""Invest in crypto!": An analysis of investment scam advertisements found in Bitcointalk". In: *APWG symposium on electronic crime research (eCrime).* IEEE. 2022, pp. 1–12.

[50] Gilberto Atondo Siu and Alice Hutchings. ""Get a higher return on your savings!": Comparing adverts for cryptocurrency investment scams across platforms". In: *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).* IEEE. 2023, pp. 158–169.

[51] Yazan Boshmaf et al. "Investigating MMM Ponzi scheme on bitcoin". In: *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security.* 2020, pp. 519–530.

[52] Massimo Bartoletti et al. "Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact". In: *Future Generation Computer Systems* 102 (2020), pp. 259–277.

[53] Weili Chen et al. "Detecting ponzi schemes on ethereum: Towards healthier blockchain technology". In: *Proceedings of the world wide web conference.* 2018, pp. 1409–1418.

[54] Weili Chen et al. "Exploiting blockchain data to detect smart ponzi schemes on ethereum". In: *IEEE Access* 7 (2019), pp. 37575–37586.

[55] Eunjin Jung et al. "Data mining-based ethereum fraud detection". In: *IEEE international conference on blockchain (Blockchain).* IEEE. 2019, pp. 266–273.

[56] Enze Liu et al. *Give and Take: An End-To-End Investigation of Giveaway Scam Conversion Rates.* 2024. arXiv: 2405.09757 [cs.CR]. URL: https://arxiv.org/abs/2405.09757.

[57] Kai Li, Darren Lee, and Shixuan Guan. "Understanding the Cryptocurrency Free Giveaway Scam Disseminated on Twitter Lists". In: *IEEE International Conference on Blockchain (Blockchain).* 2023, pp. 9–16. DOI: 10.1109/Blockchain60715.2023.00012.

[58] Anirudh Dhawan and Tālis J Putniņš. "A new wolf in town? Pump-and-dump manipulation in cryptocurrency markets". In: *Review of Finance* 27.3 (2023), pp. 935–975.

[59] Jiahua Xu and Benjamin Livshits. "The anatomy of a cryptocurrency {Pump-and-Dump} scheme". In: *28th USENIX Security Symposium (USENIX Security)*. 2019, pp. 1609–1625.
[60] Dipanjan Das et al. "Understanding security issues in the NFT ecosystem". In: *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. 2022, pp. 667–681.
[61] Muhammad Muzammil et al. "Panning for gold.eth: Understanding and Analyzing ENS Domain Dropcatching". In: *Proceedings of the Internet Measurement Conference (IMC)*. 2024.
[62] Muhammad Muzammil et al. "Typosquatting 3.0: Characterizing Squatting in Blockchain Naming Systems". In: *Proceedings of the Symposium on Electronic Crime Research (eCrime)*. 2024.
[63] Masarah Paquet-Clouston et al. "Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem". In: *Proceedings of the 1st ACM conference on advances in financial technologies*. 2019, pp. 76–88.

A Ethics

We did not interact with real users (benign or malicious) during the course of this work. We refrained from reporting identified scam websites to their respective hosting providers, as intervening would interfere with our ability to observe and understand the full scope of how cryptocurrency investment scams operate. We will be open-sourcing our code and data upon publication of this paper, so that hosting providers and operators of blocklists can take full advantage of our work.

B LLM Prompt

We use the following LLM prompt to classify scam/non-scam images (In the case of the GPT-4 vision model, we change the word 'text' to 'image'): "You are a financial advisor programmed to respond in JSON format. Your responses are limited to 'yes' or 'no', represented by the 'answer' key, and you must provide a one-word reasoning for your decision under the 'reason' key. Be sure of your answer. Determine if the provided text likely originates from a cryptocurrency investment scam website, characterized by promises of high returns from cryptocurrency investments. If the text suggests a low probability of being a scam, does not prompt users to log in/sign up, seems like a news site, or does not solicit users to contact the site for investing in cryptocurrency, respond 'no'."

C Scam Website Analysis: Additional Insights

C.1 Domain Names

Table 6: Top 5 most frequent TLDs among investment scam websites

TLD	Detected Scams	Tranco Top 10K
com	27,955 (64%)	5,101 (51%)
net	2,394 (5%)	456 (5%)
org	2,079 (5%)	538 (5%)
online	1,264 (3%)	6 (<1%)
ltd	1,012 (3%)	0 (0%)
(Total)	*34,704 (80%)*	*6,101 (59%)*

Table 7: Top 10 words in investment scam domain names detected by Crimson

Word	Detected Scams	Tranco Top 10K
trade	9,706 (22%)	7 (<1%)
crypto	4,458 (10%)	2 (<1%)
fx	4,078 (9%)	2 (<1%)
invest	4,025 (9%)	0 (0%)
coin	3,181 (7%)	4 (<1%)
capital	2,926 (6%)	4 (<1%)
bit	2,664 (6%)	24 (<1%)
global	2,249 (5%)	21 (<1%)
bitcoin	1,776 (4%)	3 (<1%)
mine	1,551 (4%)	0 (0%)
(Total)	*71%*	*<1%*

C.2 Web Design-based Clustering

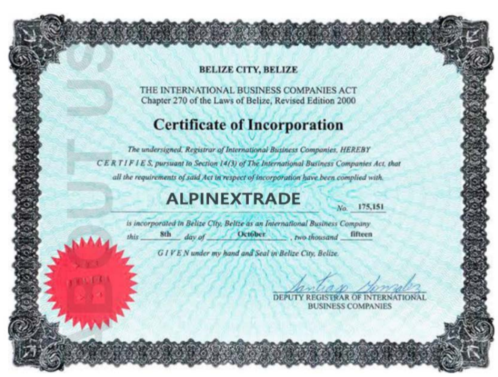
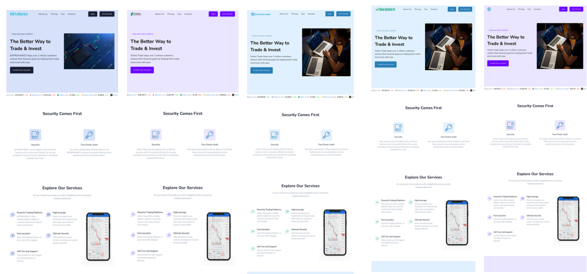
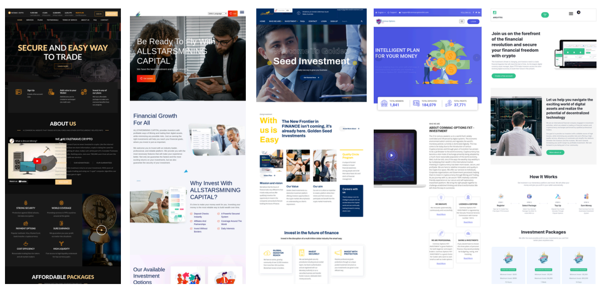


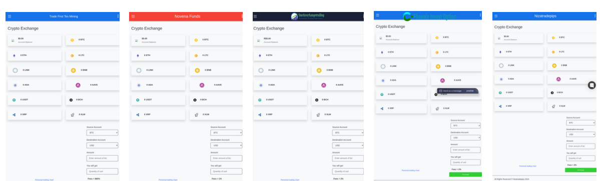
Figure 6: Fake certificate displayed by the investment scam website `alpinextrade.com` *on their home-page*



(a) Example clusters of investment scam websites that have similar design.



(b) Example clusters of investment scam websites that have dis-similar design.



(c) Example clusters of investment scam website deposit pages that have similar design.

Figure 7: Example clusters of investment scam websites, categorized by design similarities, dissimilarities, and deposit page designs.

C.3 IoC-based Clustering

Table 8: YouTube channels embedded within the HTML of detected cryptocurrency investment scam websites, along with their subscriber counts, channel titles, example video titles, and the number of scam websites that linked to each channel.

Channel ID	# Subscribers	Channel Title	Example Video Title	# Scam Websites
UCmhqA0PXvpSj8kuN3zshpDw	6.2K	InstaForex Official	Forex forecast 10/10/2024: EUR/USD, USDX, Gold and Bitcoin	24
UC8ATbwlPxMCrqFlYweaizgA	Blocked	N/A	N/A	23
UCRF2-5W_uwflhpj6Hf6r4Jw	32K	Intelligent Cryptocurrency - Dirk Crypto Diggy	How to (Realistically) Make 100k From Crypto in 2024	16
UCvBNXhZD6XlMCb9FY9KJxqw	Blocked	N/A	N/A	12
UCzH0C03Gy8uHyKr-Y59cwJg	3K	PrimeXBT	Trade on an easy-to-use platform - PrimeXBT	12
UCBEu2buLFT8nX2ffFV76XNQ	134	101financial.app - A New Social Way to Invest	EASY WAY TO DEPOSIT BALANCE IN 101.INVEST!	11
UCHxfIung8_7Z7zCk6NTaYrQ	Blocked	N/A	N/A	10

Table 9: Instagram account usernames embedded within the HTML of detected cryptocurrency investment scam websites, along with the number of scam websites that linked to each username.

Instagram Username	# Scam Websites
hyiprio	185
zeus.strategy	31
cryptotabme	29
pangmancapital	28
miningautomatic	26

Table 10: Telegram handles embedded within the HTML of detected cryptocurrency investment scam websites, along with the number of scam websites that linked to each handle.

Telegram Handle	# Scam Websites
flexytrading1	189
PhoenixFX	130
bitcoinminetrix	67
CryptoTabChannel	29
klassiccapitalchannel	23

D Keyword Filters

Table 11: Keyword filters used in the URL-filter and Content-filter modules

#	URL-filter	Content-Filter		
		Invest Words	Coin Words	Context Words
1	crypto	invest	cryptocurrency	deposit
2	fx		crypto	withdraw
3	earn		bitcoin	reward
4	deposit		ethereum	growth
5	trade		cardano	gain
6	capital		ripple	capital
7	invest		binance	potential
8	global		shiba inu	wallet
9	bit		dogecoin	safe
10	mining		solana	secure
11	ltd		tether	fund
12	finance		tron	profit
13	trade		polkadot	insurance
14	miner		eth	wealth
15	trust		btc	send
16	profit		xrp	transfer
17	asset		ada	sell
18	cardano		bnb	buy
19	funding		shib	trade
20	capitals		doge	asset
21	fund		sol	client
22	limited		usdt	solution
23	chain		trx	funding
24	digital		dot	
25	btc		algo	
26	assets		litecoin	
27	wealth		chainlink	
28	coin		uniswap	
29	option		pancakeswap	
30	prime		avalanche	
31	bitcoin		neo	
32	exchange		iota	
33	money		aave	
34	eth		luna	
35	ethereum		synthetix	
36	cryptocurrency		theta	
37			grt	
38			1inch	
39			sushi	
40			matic	
41			btcusd	
42			usdbtc	
43			ethusd	
44			usdeth	
45			adausd	
46			usdada	
47			xrpusd	
48			usdxrp	
49			bnbusd	
50			usdbnb	
51			shibusd	
52			usdshib	
53			dogeusd	
54			usddoge	
55			solusd	
56			usdsol	
57			usdtusd	
58			usdusdt	

E Financial Losses: Additional Insights

E.1 Bitcoin

Table 12: Top 20 investment scam domains sorted by revenue earned through Bitcoin payments, along with their Bitcoin wallet addresses, revenue in USD, and whether they were active at the time of writing. Web browsers show no deception warnings when a user visits any of these websites. Almost all of the listed websites are active, and are still potentially receiving payments from unsuspecting users

Scam Domain	Wallet Address	Revenue	Active?
capitalfxfinance.org	bc1q2em...tcuet	257K	✓
bitgainscapital.com	bc1qtut...tnnn9	108K	✓
duxtonroztrade.com	bc1qwzs...ywx00	86K	✓
digitechmininghub.com	bc1qnzn...xcnck	51K	✗
coincipher.co	bc1qhy9...uzucwn	50K	✓
tslasafeinvest.com	bc1qlcq...sqds6r5	49K	✗
cryptomineenergy.com	bc1qfgx...xxjlrk	45K	✓
crudeportlimited.com	bc1q5c6...lak0z	44K	✓
growthmatrixinvestment.com	bc1q7gn...hapwzm	33K	✓
finance-extra.com, hextechcryptofarm.com	bc1qarp...e8kuj9	30K	✓,✓
tradesprofitly.com	bc1qy33...5z09e	29K	✓
apextradexf.com	bc1qpet...ndqltu	24K	✓
capitalwheelinvestmentcompany.com.ld-bnk.com	bc1q3q7...mh6hqw	22K	✓
compasscloudminings.com	bc1q5g4...3vtykp	22K	✓
bridgefastltd.com	bc1q6mm...skskyq	21K	✓
horizoncrypto.ltd, coinfarmlandltd.com	bc1quxl...6amrrt	21K	✗,✓
xtradeconnect.com	bc1q4gc...yalx77y	21K	✓
mytradepay.com	bc1qxku...apk8kgp	18K	✓
primepinnaclepurse.com	bc1qpcq...vgw8hz	17K	✓
forcasttradeslimited.com	bc1qnr3...uw7jle	17K	✓

E.2 Ethereum

Table 13: Top 20 investment scam domains sorted by revenue through Ethereum payments detected by Crimson, along with their Ethereum wallet addresses, revenue in USD, and whether they were active at the time of writing.

Scam Domain	Wallet Address	Revenue	Active?
apexasset-management.net	0x8507...7dd66	193K	✓
brclimited.com	0x7c53...f950	35K	✓
tradepeakinvest.com	0x1991...26a2	18K	✓
horizoncrypto.ltd, findexglobalchain.com	0xbd95...a45d	14K	✗ ✓
kings-investment.co	0x68e7...ff4c	10K	✗
aqrisprime.io	0xecd2...88d3	9K	✗
firstclassrader.mdxspacetrade.com	0x704b...5ae9	9K	✓
exgrowfundlimited.com	0xdda9...98e2	7K	✓
capitalprimextrades.com	0x0ec2...10a4	7K	✓
bi-investments.com	0x2428...afa8	7K	✓ ✓
equityegdecapital.com	0x2931...ec37	6K	✓
mytradepay.com	0xc583...e532	6K	✓
tfxprimes.com	0x28e9...8159	5K	✓
renoexperttrade.online, bitcorefxtrade.com, falconexchangealtcoin.online, fxtradingtrust.com	0xb664...3ede	4K	✓ ✓ ✓ ✓
cryptospacecapitals.com, horizoncrypto.ltd	0xb4eb...5ac1	2K	✓ ✗
globalsprovest.com, globalsprovest.com.shipscago.com	0xd18e...58bd	2K	✓ ✓
ventures-fundsfx.com	0x0571...9f30	2K	✗
goldcoinridge.com	0x2452...c4f6	2K	✓
graceautoinvestsltd.com	0x0633...a39a	2K	✗
e-musktrading.com	0x11d9...0c37	1K	✓