Published: 13 December 2024

RESEARCH-ARTICLE

Piecing Together the Jigsaw Puzzle of Transactions on Heterogeneous Blockchain Networks

XIAOHUI HU, Huazhong University of Science and Technology, Wuhan, Hubei, China

HANG FENG, Zhejiang University, Hangzhou, Zhejiang, China

PENGCHENG XIA, Huazhong University of Science and Technology, Wuhan, Hubei, China

GARETH TYSON, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

LEI WU, Zhejiang University, Hangzhou, Zhejiang, China

YAJIN ZHOU, Zhejiang University, Hangzhou, Zhejiang, China

View all

Proceedings of the ACM on Measurement and Analysis of Computing Systems, Volume 8, Issue 3 (December 2024)
https://doi.org/10.1145/3700424
EISSN: 2476-1249

Piecing Together the Jigsaw Puzzle of Transactions on Heterogeneous Blockchain Networks

XIAOHUI HU[†], Huazhong University of Science and Technology, China
HANG FENG, Zhejiang Univeristy, China
PENGCHENG XIA[†], Huazhong University of Science and Technology, China
GARETH TYSON, Hong Kong University of Science and Technology (GZ), China
LEI WU, Zhejiang Univeristy, China
YAJIN ZHOU, Zhejiang Univeristy, China
HAOYU WANG[*†], Huazhong University of Science and Technology, China

The Web3 ecosystem is increasingly evolving to multi-chain, with decentralized applications (dApps) distributing across different blockchains, which drives the need for cross-chain bridges for blockchain interoperability. However, it further opens new attack surfaces, and media outlets have reported serious attacks related to cross-chain bridges. Nevertheless, few prior research studies have studied cross-chain bridges and their related transactions, especially from a security perspective. To fill the void, this paper presents the first comprehensive analysis of cross-chain transactions. We first make efforts to create by far the largest cross-chain transaction dataset based on semantic analysis of popular cross-chain bridges, covering 13 decentralized bridges and 7 representative blockchains, with over 80 million transactions in total. Based on this comprehensive dataset, we present the landscape of cross-chain transactions from angles including token usage, user profile and the purposes of transactions, etc. We further observe that cross-chain bridges can be abused for malicious/aggressive purposes, thus we design an automated detector and deploy it in the wild to flag misbehaviors from millions of cross-chain transactions. We have identified hundreds of abnormal transactions related to exploits and arbitrages, etc. Our research underscores the prevalence of cross-chain ecosystems, unveils their characteristics, and proposes an effective detector for pinpointing security threats.

CCS Concepts: • **Security and privacy** → *Software and application security*.

Additional Key Words and Phrases: Cross-chain, Decentralized Finance, Blockchain, Transaction Analysis

ACM Reference Format:
Xiaohui Hu, Hang Feng, Pengcheng Xia, Gareth Tyson, Lei Wu, Yajin Zhou, and Haoyu Wang. 2024. Piecing Together the Jigsaw Puzzle of Transactions on Heterogeneous Blockchain Networks. *Proc. ACM Meas. Anal. Comput. Syst.* 8, 3, Article 42 (December 2024), 27 pages. https://doi.org/10.1145/3700424

[*]Corresponding Author: Haoyu Wang (haoyuwang@hust.edu.cn).
[†]The full name of the authors' affiliation is Hubei Key Laboratory of Distributed System Security, Hubei Engineering Research Center on Big Data Security, School of Cyber Science and Engineering, Huazhong University of Science and Technology.

Authors' Contact Information: Xiaohui Hu, Huazhong University of Science and Technology, Wuhan, China, xiaohui_hu@hust.edu.cn; Hang Feng, Zhejiang Univeristy, Hangzhou, China, h_feng@zju.edu.cn; Pengcheng Xia, Huazhong University of Science and Technology, Wuhan, China, xpc357@hust.edu.cn; Gareth Tyson, Hong Kong University of Science and Technology (GZ), Guangzhou, China, gtyson@ust.hk; Lei Wu, Zhejiang Univeristy, Hangzhou, China, lei_wu@zju.edu.cn; Yajin Zhou, Zhejiang Univeristy, Hangzhou, China, yajin_zhou@zju.edu.cn; Haoyu Wang, Huazhong University of Science and Technology, Wuhan, China, haoyuwang@hust.edu.cn.

ACM 2476-1249/2024/12-ART42
https://doi.org/10.1145/3700424

1 Introduction

Since the inception of Bitcoin [60], blockchain has garnered significant attention and seen substantial development. Numerous different blockchains have been established to fulfill diverse demands within the digital currency ecosystem. Bitcoin and Namecoin [36] represent the first-generation blockchains, primarily capable of executing token transfers and locks. The second generation of blockchains, represented by Ethereum [23], was then introduced to enhance the versatility of transactions and allow complex logic across various scenarios. For example, through Ethereum, it is possible for developers to develop Decentralized Applications (dApps) based on smart contracts. This has stimulated the rapid development of Decentralized Finance (DeFi), going from $700 million in total value locked (TVL) at the start of 2020 to a peak of $109.01 billion on May 27, 2024 [24].

This proliferation of blockchain networks has created serious challenges for interoperability though, preventing seamless communication and asset transfers. This has naturally posed a hindrance to the advancement of DeFi more broadly, as a DeFi project may aspire to maintain a presence on multiple chains. A user may also opt to move assets across blockchains, to evade high gas fees on congested networks. To support this, a number of cross-chain projects have been developed to enable interoperability between blockchains. Cross-chain technologies can be generally classified into three categories [34], *i.e.*, notary schemes, sidechains, and hash-locking.

The cross-chain ecosystem is vast, yet fraught with intricate issues from organization to security. As of now, there are at least 56 decentralized cross-chain bridges (*e.g.*, Stargate[20], Arbitrum Bridge[7]) facilitating asset transfers across different blockchain networks. These bridges collectively handle a daily volume exceeding 251 million, comprising 6% of the volume in DeFi thereby fostering a vibrant digital currency ecosystem [10]. Unfortunately, this vast ecosystem also expands the threat landscape. According to [13], the global Web3 ecosystem suffered over $4 billion loss due to exploits in 2022, with cross-chain bridges accounting for over 33.32%. As of July 2024, the cumulative lost and recovered funds within the cross-chain ecosystem have exceeded $2.39 billion [15]. A previous study has claimed that malicious users exploit bridges to transfer stolen funds to other blockchains [49], which not only poses challenges for the automated tracking of stolen funds but implies a complexity of cross-chain user composition. Moreover, the lack of a unified implementation mechanism for cross-chain bridges hinders understanding of their different processes and contributes to the proliferation of diverse bridge tokens.

Existing studies lack an in-depth understanding of the cross-chain ecosystem. The absence of datasets and challenges in collecting cross-chain data hinder both comprehensive research into the cross-chain ecosystem and the automated tracking of bridged funds. Most works are confined to cross-chain technologies, presenting categorization, development [72], and summary [33, 48, 57, 66]. To mitigate security issues, Zhang *et al.* [77] categorized bugs in the cross-chain system into three categories and introduced Xscope for security violation detection. To the best of our knowledge, no prior work has conducted systematic measurements of cross-chain transaction volumes, token usage, and user composition. This is largely because the paucity of empirical data makes it arduous to track the destination and distribution of bridged funds on the target chain.

However, it is non-trivial to address these challenges in investigating cross-chain transactions. Collecting transactions of various cross-chain bridges is a demanding endeavor, primarily due to the limited availability of bridges supporting retrievals of historical transactions. Moreover, the varied semantics of bridge contracts make it impossible to filter out different bridge transactions in a unified way. Additionally, we have identified obvious inaccuracies in third-party platforms that present statistical data in dashboards, which could mislead the community.[1] From a general perspective, major challenges include *(i)* how to build a cross-chain dataset while ensuring its

[1]For example, in data provided by DefiLlama [10], the cross-chain transaction volume from June 21st to 23rd is zero.

integrity and accuracy; *(ii)* how to calculate the dollar value of cross-chain funds and identify the identities of involved addresses when building fund flow graphs; and *(iii)* how to obtain the token price at the time of a transaction thereby sensitively detect misbehaviors.

Our work. To address these challenges, we conduct the first comprehensive study of the cross-chain ecosystem. Our data collection pipeline is composed of multiple phases (§3). The first phase is to uncover cross-chain transaction pairs by observing the cross-chain patterns, identifying specific functions and events for each bridge, and then taking a *Tx_in* as input to match it with the corresponding *Tx_out*. Then, we build our large-scale dataset based on complete matching transaction pairs derived from phase 1. The third phase takes advantage of some external resources (*e.g.*, some cross-chain transactions are retrievable through APIs) for validating and complementing our dataset. In this way, we have compiled a comprehensive dataset that ensures both accuracy and integrity. Our final dataset covers the period from March 1st, 2021 to January 1st, 2024, encompassing a total of 80,046,355 cross-chain transactions from 13 bridges on seven chains.

Based on the dataset, we characterize the nature of cross-chain transactions (§4). We find that Stargate stands out among bridges with 67% of transaction volume. We further determine that stablecoins (*i.e.*, USDC and USDT) are the most popular tokenized assets in this ecosystem. We then proceed to study potential security issues. We identify 576 malicious accounts involved in 1,929 cross-chain transactions, transferring 21,383 Ether and other ERC20 tokens. This finding motivates us to track fund flow. For this, we construct fund flow graphs to explore the distribution of cross-chain funds (§5.3). We show that 24.42% of cross-chain transactions flow to exchanges. Taking advantage of this, we track funds lost in attacks on DeFi projects during January 2024 (§5.5). We further observe that cross-chain bridges can be abused for malicious/aggressive purposes. Thus we design an automated detector and deploy it in the wild to flag misbehaviors (including exploits and arbitrages) from millions of cross-chain transactions. We also discover and report severe data omissions and errors in ChainBase [1], resulting in bug fixes.

Contribution. In short, we make the following contributions.

- **Large-scale dataset.** We build by far the largest cross-chain transaction dataset for our community. We make efforts to ensure the accuracy and integrity of the dataset, which will be released to the community for further research.
- **In-depth study of cross-chain transactions.** We conduct the first in-depth study of the cross-chain ecosystem, exploring its prevalence in terms of token and user compositions. We further track the flow of cross-chain transactions to accurately depict their targets and distribution.
- **Automated alerting system.** We summarize the patterns of four kinds of behaviors related to cross-chain bridges, and create an automated detection system called CrossAlert. CrossAlert is deployed in the wild, and works as a whistle-blower for cross-chain bridges. We have identified hundreds of abnormal transactions related to exploits and arbitrages.

2 Background

2.1 Transactions and Layer Services in Blockchains

Transactions. Transactions are signed messages sent from one account to another, recording state changes of accounts. Transactions sent by EOAs are called external transactions, which can be directly obtained by parsing blocks, while internal transactions are triggered by external transactions and cannot be obtained directly.

Layer Services. A Layer 1 blockchain refers to the foundational network that defines the decentralized consensus mechanism. Examples include Bitcoin, Ethereum, and Solana [28]. These

blockchains are crucial for providing the fundamental infrastructure to decentralize interactions. However, scalability is a critical concern of these Layer-1 blockchain technologies [68]. Layer 2 blockchains strive to enhance transaction processing speeds [39, 59, 61], reduce latency, and lower fees [46]. To achieve this, Layer 2 approaches process transactions off-chain, thereby reducing the load on the blockchain and significantly increasing transaction speed. Existing works [44, 64, 69] have demonstrated that implementing a Layer 2 solution significantly reduces gas fees, thereby making DeFi activities more accessible to a broader audience.

2.2 Decentralized Finance (DeFi) Tokens

DeFi tokens, issued on blockchains, represent a diverse set of cryptocurrencies and function through smart contracts. These tokens grant users access to a suite of financial applications and services built on the blockchain.

Wrapped Tokens. Wrapped crypto tokens are cryptocurrencies pegged to the value of another underlying crypto or assets like gold and stocks that are utilized within Decentralized Finance (DeFi) platforms. These tokens are supported in a 1:1 exchange ratio by the underlying token and generally follow a deposit-mint-withdraw approach. In this paradigm, a user deposits tokens on a specified address on a blockchain and then, thanks to cryptographic methods, a corresponding token is created (minted) in reverse [31].

Cross-chain Tokens. Cross-chain tokenized assets are tokens that can be transferred and traded across different blockchains. They are accessible on multiple chains, either as wrapped assets or as tokens natively minted on chains. Some bridges issue their tokens to facilitate authentication for a broader range of tokens. For example, Multichain [18] issues the anyUSDC token, which can securely flow between blockchains thanks to the burn and mint cross-chain mechanism, and be exchanged for USDC. We compute the market value of these cross-chain wrapped tokens by leveraging their underlying tokens.

2.3 Cross-chain Techniques

Cross-chain technology tries to build a bridge of trust between chains, preventing different blockchains from being isolated islands, and realizing asset interoperability between chains [62].

Cross-chain Bridges. Pau *et al.* have classified bridges into four different types, namely centralized, somewhat centralized, decentralized, and untrusted bridges [40]. They also outline the characteristics and key considerations of bridges. These include custody, relayer, verification, incentive, and asset-transferring methods. Our research focuses on decentralized cross-chain bridges. These bridges typically leverage smart contracts to lock or burn assets on the source chain while issuing or minting corresponding assets on the target chain. Based on daily total volume ranking, notable decentralized cross-chain bridges include *Stargate, Arbitrum Bridge, Hop, Synapse, Celer cBridge, and Multichain.*

Cross-chain Transactions. As illustrated in Figure 1, the communication between blockchains involves two blockchains: the source and the target [33]. On the source chain, a depositor initiates a cross-out transaction (*Tx_out*) and their assets are burned or locked by the bridge contract. While on the target chain, the corresponding cross-in transaction (*Tx_in*) is generated and after verification, a relayer signs *Tx_in* so that assets will be released to the recipient on the target chain. Only when these two transactions are matched, a complete cross-chain transaction pair is determined.

3 Data Collection and Study Design

In this section, we present our general process for identifying cross-chain bridge transactions, as illustrated in Figure 2. It is challenging to ensure data integrity and matching accuracy across

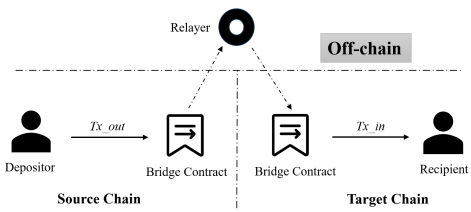


Fig. 1. Architecture of top bridges.

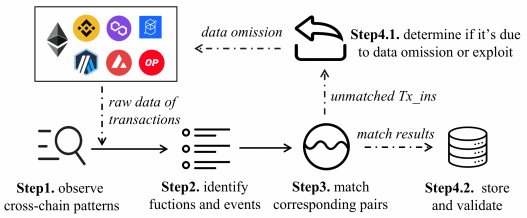


Fig. 2. Data collection process.

different chains. Thus, we share the insights gleaned through the process, before introducing our research questions.

3.1 Identifying Cross-chain Transactions

To the best of our knowledge, there are no existing public datasets aggregating transactions of multiple cross-chain bridges, nor are there existing approaches to identify cross-chain transactions automatically. Therefore, we develop a methodology based on contract semantics to extract cross-chain transactions from 13 representative cross-chain bridges,[2] allowing us to construct a comprehensive dataset.

Identifying functions and events. For each decentralized cross-chain bridge, we first seek to understand its contract semantics to identify its cross-chain-related functions and events. As described in Section 2.3, a complete cross-chain procedure involves a cross-out transaction (*Tx_out*) on the source chain and a cross-in transaction (*Tx_in*) on the target chain. Typically, transactions on both chains carry essential information (described in Appendix Table 7) including source/target chains, from/to addresses, token contract addresses, and amounts. To extract these key fields, a prerequisite is to identify the function calls and emitted events that contain essential information in a cross-chain contract. However, one challenge is that it is possible that a bridge deploys multiple contracts and, within a single contract, there may be various function entries for cross-chain transactions. For example, Synapse invokes `withdraw` to transfer ETH while invokes `withdrawAndRemove` to transfer ERC20 tokens on the Ethereum. Thus, it has been necessary to perform a deep analysis of these contracts to understand each cross-chain bridge's operations in detail. Specifically, we need to expand the scope of semantic analysis to include source code and all deployed contracts for each bridge. We must also combine the data from cross-chain-related event logs and function calls within one transaction.

Extracting on-chain activities. Next, using the signatures of the identified functions and events, we can filter out all cross-chain transactions per-chain. Further, semantic analysis of bridge contracts assists us in parsing essential cross-chain information from the bytecode, which is directly retrievable from blockchains. We note some special transactions of Celer cBridge should be excluded. These transactions use the *Tx_in* function to execute refunds for failed *Tx_out* and they are not strictly cross-chain transactions. In such cases, paired *Tx_out* and *Tx_in* are on the same chain, and the sender address for the *Tx_in* is fixed, therefore we can identify and exclude them.

Constructing the collector. We have embedded the above functionality into our data collection tool. We have deployed our collector across seven representative blockchains[3] with the highest total value locked (TVL). Atop of this, we collect cross-chain transactions on 13 bridges. Using

[2]Stargate, Arbitrum Bridge, Optimism Bridge, Polygon Pos Bridge, Avalanche Bridge, Across, Hop, Synapse, Celer cBridge, Multichain, PolyNetwork, SwitcheoNetwork, WormHole

[3]Ethereum, BSC, Polygon, Arbitrum, Optimism, Avalanche, and Fantom

Chainbase [1] or blockchain explorers [12] can significantly simplify the process of collecting and parsing data, as they support the direct retrieval of complete transaction data using identified cross-chain-related signatures. In contrast, using Geth [2] requires multiple requests to the node, which can lead to network congestion and significant data delays, regardless of whether using our own Geth nodes or third-party services. Additionally, maintaining Geth nodes on seven blockchains incurs prohibitively high storage costs. *That said, it is challenging to ensure data integrity when using these explorers.* For example, issues such as network request failures and API limits can lead to data omissions. Indeed, our experience indicates that, when the frequency of requests for the Geth nodes is too high, the results may exhibit missing or erroneous data. Despite this, on balance, we find that blockchain explorers offer the best trade-off. To identify erroneous cases, when a *Tx_in* cannot be matched, we calculate the corresponding block range on the source chain and initiate a repetitive collection with Geth nodes. Note, in this process, we also discovered significant data inaccuracy in Chainbase and reported it, receiving confirmation from their team.

3.2 Matching Transaction Pairs

The previous step allows us to extract all cross-chain transactions. Next, it is necessary to match these transactions across the chains (*i.e.* link a transaction on one chain with a transaction on another). This is challenging to automate with accuracy. For example, a *Tx_out* may remain in pending status or encounter a failure due to various reasons (such as network congestion, insufficient liquidity, and minimal amounts). Consequently, its corresponding *Tx_in* will not be triggered on the target chain. However, each *Tx_in* is expected to be matched with a specific *Tx_out* because the release of assets must occur only after confirming that the assets are locked or burned on the source chain. In our design, the matching process therefore takes each *Tx_in* as input and seeks its corresponding *Tx_out* in the remaining data. We rely on two methods to match the transactions, detailed below.

Matching transactions with unique values. The first method relies on matching uniquely identifiable values within the transactions. Most decentralized bridges[4] are either designed or can be parsed based on open-source code to extract a specific field in related functions or events. The values of certain fields in each transaction are unique (*e.g.*, regarding a *Tx_in* of Multichain, 'srcTxHash' is recorded in the cross-in event 'LogAnySwapIn', representing the corresponding transaction hash on the source chain). Therefore, a *Tx_in* and a *Tx_out* with the same field value can be directly matched as a pair.

Matching transactions without unique values. Some bridges are also designed where fields are conditionally repetitive (*e.g.*, the 'nonce' field of Across[5] and the 'txid' of Hop [14]), and other bridges may even operate without any observable unique field. To overcome this, we propose a novel heuristic algorithm to automatically search for the corresponding *Tx_out* for each *Tx_in*on these bridges. First, given the existence of a conditionally repetitive field, we take all unmatched *Tx_out*s with the same filed value as candidates. While for a *Tx_in* without any specific field, we identify candidates using restrictive information (*e.g.*, Arbitrum Bridge *Tx_out* specifies the replayer address on the target chain). Second, we then narrow down the candidates based on the consistency of information including the source/target chain, depositor, recipient, and bridged token. We also check that the transferred amount in *Tx_in* is no larger than that in *Tx_out*, and the timestamp of *Tx_in* must be no earlier than that of *Tx_out*.

Chronological matching is done only when multiple transactions share the same information and have an equal number of *Tx_in*s and *Tx_out*s. If not, we suspect data omissions and abnormal matches. In such cases, we re-collect and re-match transactions within the block range, manually

[4]Across, Avalanche Bridge, Celer cBridge, Hop, Optimism Bridge, WormHole, Multichain, PolyNetwork, Synapse, Stargate

verifying any unmatched ones. Empirical data shows that fewer than 0.002% of transactions share identical information except for the time.

3.3 Validating Results

Our collector continuously monitors and filters out real-time cross-chain transactions across the seven blockchains in parallel. As of January 1st, 2024, we have amassed 80,046,355 transactions since March 2021, involving 3,413,640 unique accounts and 8,701 different token contracts. After matching work, 99.8% transactions are successfully matched. However, the absence of publicly available ground truth data poses significant challenges for data validation. Thus, we take several steps to validate and complement our dataset.

Validating matching accuracy. It has to be noted that our data collection is limited to the seven specific chains covered. Thus, the matched transactions pertain exclusively to those (where both the source and target chains are among these seven). Three methods are employed to validate our matching accuracy. *(i)* We exploit the APIs available for some bridges. Hop, Celer cBridge, and WormHole all provide APIs to allow users to obtain ground truth cross-chain transactions for a specific token. These APIs allow us to retrieve detailed ground truth information about a transaction including the corresponding one on the remote target chain. *(ii)* We randomly sample and manually inspect 3,000 matched transactions from bridges that do not support transaction retrievals. Subsequently, we verify the cross-chain information for all matched pairs and the results are completely consistent. *(iii)* We check unmatched *Tx_in*s within the matching scope, which requires their source chain to be in the seven-chain range, and determine whether the root cause is a flaw in the matching algorithm. Results show that none of them resulted from incorrect matching, indicating the correctness of our matching algorithm.

Validating data integrity. It is compelling to validate data integrity with matching results. We manually check all unmatched *Tx_in*s to determine whether they stem from data omission. The validation results show that of the 0.2% unmatched transactions, 99.96% were due to data omission, indicating that the rate of data omission is approximately 0.2%. We subsequently rescan corresponding block ranges to collect missing items, then repeatedly match and validate results. As a result, 94 *Tx_in*s are abnormal, which we introduce in §6.3.1. All others are successfully matched.

Since only the WormHole API allows us to obtain all transactions, we query it for validation. We confirm that all WormHole transactions having their source/target chain within the seven-chain range are recorded correctly in our dataset. Consequently, through our repeated validation and data collection efforts, we have compiled a comprehensive dataset that can ensure both accuracy and integrity.

3.4 Research Questions

Our study is driven by the following research questions (RQs):

RQ 1 ***How prevalent are cross-chain transactions in the cryptocurrency ecosystem?*** Although the cross-chain ecosystem is sometimes reported in the media, there is still a lack of analysis pertaining to their transaction scale, development, and related tokens. Therefore, it is necessary to investigate: RQ1.1) *What are the volumes of transactions across the different cross-chain bridges and how do volumes change?* Considering the variety of cross-chain bridges, we focus on bridges with a relatively high volume. RQ1.2) *Which tokens are mainly used?* Analyzing related tokens can help us calculate the value, and understand the overall scale of asset circulation in the ecosystem. And RQ1.3) *What are the compositions and behavioral characteristics of users?*

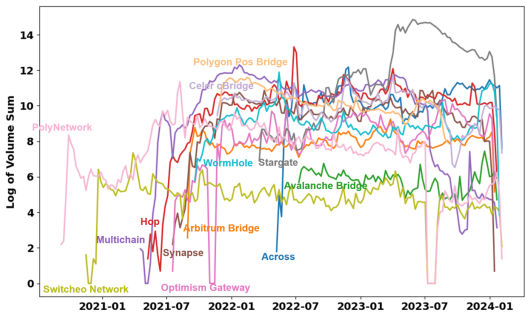


Fig. 3. Daily transaction volumes of 13 bridges.

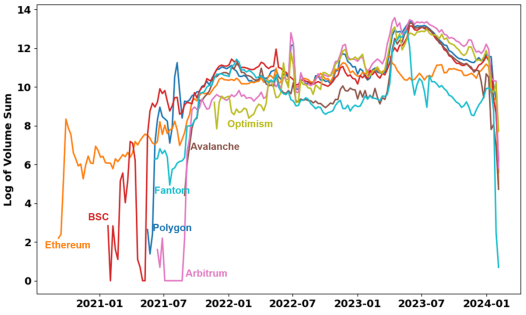


Fig. 4. Daily transaction volumes of seven chains.

RQ 2 ***What are the target destinations and distribution of cross-chain funds on the target chain?*** We investigate the termination of cross-chain funds on the target chain, thereby incurring user potential purposes for bridging funds. For this, we construct complete fund flow graphs for *Tx_in*s and observe the distribution of funds across chains. In addition to tracking funds, our fund flow graphs also enable us to explore behaviors of malicious accounts and investigate characteristics of common users.

RQ 3 ***What misbehaviors can be observed and detected in the cross-chain ecosystem?*** We explore exploits, arbitrage, and other abnormal cases that result in illicit profits, focusing on the identification of anomalies to build an early warning system.

4 Characterizing Cross-chain Transactions

Based on collected transactions from 13 decentralized cross-chain bridges on seven blockchains, we conduct a comprehensive investigation to explore *transaction volume, token usage* and the *composition of users* to discover the popularity of cross-chain ecosystem.

4.1 Transaction Volume

We first investigate cross-chain transaction volumes on different bridges and blockchains. Our goal is to quantify the prevalence of cross-chain usage. We observe a significant volume of cross-chain transactions. During the period from March 1st, 2021 to January 1st, 2024, there are **over 80 million (80,046,355)** transactions from 13 cross-chain bridge projects on seven blockchains. This consists of 43,270,084 cross-out transactions (*Tx_out*) and 36, 777, 775 cross-in transactions (*Tx_in*). **3,413,640 distinct accounts** are found to have completed cross-chain transactions and there are **8,701 token contracts** involved in cross-chain transfers. Of particular note is that, from April 2023 to January 2024, the transaction volume of these 13 bridges accounts for approximately 4.8% of the 1.47 billion transactions across the seven chains. Figure 3 presents the daily volumes of 13 bridges. To further investigate the prevalence, we demonstrate significant observations, and next analyze the transaction distribution on different bridges and blockchains.

4.1.1 Significant Observations. We observe significant changes in cross-chain transaction volume, which are illustrated in Figure 3, and then collect reports to clarify the root cause. Most bridges exhibit a gradual increase in usage starting from early 2022, peaking in 2023, and showing a slight decline by early 2024. This indicates that the cross-chain services have been progressively adopted and utilized by more users over the past two years. The volume growth in July 2022 and April 2023 could be linked to airdrops facilitated by Hop and Stargate, respectively. In the initial phase of the cross-chain service, Multichain attracted the largest number of users, due to its supporting

Table 1. Transactions across seven blockchains indicate their interactions.

Source \ Target	Ethereum	BSC	Polygon	Arbitrum	Avalanche	Optimism	Fantom
Ethereum	127	310K	258K	881K	112K	516K	77K
BSC	280K	59	1,927K	1,176K	1,450K	422K	460K
Polygon	278K	1,653K	54	1,736K	1,921K	863K	835K
Arbitrum	920K	1,327K	2,035K	2	1,460K	4,021K	335K
Avalanche	79K	1,271K	1,806K	1,075K	8	393K	649K
Optimism	366K	448K	852K	3,435K	464K	5K	105K
Fantom	49K	436K	728K	178K	562K	81K	11

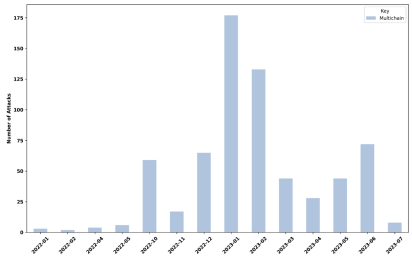


Fig. 5. Numbers of attacks on Multichain per month.

extensive variety of blockchains and tokens. However, it is evident that usage significantly declined after July 2023, as the funds in Multichain's privileged accounts were transferred.

4.1.2 Transactions on Bridges. We discover a significant imbalance of cross-chain transaction volume among the 13 bridges. This is a reflection of user preference towards different bridges. Stargate stands out as the most popular bridge, covering 67% of transactions. The following are Multichain, Hop, Across, and Celer cBridge, consisting of 9.79%, 7.08%, 3.98%, and 3.97% of the volume, respectively. Exploring the factors influencing user choices of cross-chain bridges can uncover potential user needs. To this end, we characterize different bridges as below:

- **Security.** Stargate, Hop, and Across have not experienced any known security incidents. Multichain promptly addressed vulnerabilities after being attacked in January 2022 and did not encounter significant issues until its service was shut down in July 2023. We use Figure 5 to illustrate the numbers of attacks on Multichain per month. However, PolyNetwork, WormHole, and Celer cBridge experienced several severe exploits in 2022 and their transaction volumes are significantly lower in comparison, indicating low user trust in bridges with worse security.
- **Time latency.** The time taken from *Tx_out* to the corresponding *Tx_in* is referred to as time latency. We observed transactions with a latency period of fewer than ten minutes account for 98.4% of all Celer cBridge transactions, 94.7% of Across transactions, and 80.4% of Stargate transactions. While for other bridges, including WormHole, PolyNetwork, Optimism Bridge *et al.*, the majority of latency falls within 10-30 minutes.
- **Supported tokens.** Despite Multichain and WormHole falling behind other bridges in terms of latency and security (especially with WormHole experiencing three exploits), they still attract a certain user base. This could be attributed to the diversity of their supported tokens. The one that supports the most tokens is Wormhole, followed by Multichain and ArbitrumBridge, which support 4,926, 1,487, and 981 tokens, respectively. Next is OptimismGateway, supporting 430 tokens, cBridge with 411, PolyNetwork with 367, Synapse with 124, Avalanche Bridge with 123, and Polygon POS Bridge with 103. SwitcheoNetwork supports 77 tokens. However, the chain with the largest transaction volume, Stargate, only supports 43 tokens, with only Across and Hop supporting fewer tokens, at 39 and 21, respectively.

Based on the above analysis, we can infer that high security, short latency, and a wide variety of supporting tokens are factors influencing users' choice to interact with cross-chain bridges.

4.1.3 Transactions on Blockchains. Arbitrum leads with over 11.35 million *Tx_out* and 8.44 million *Tx_in*, comprising 24.73% of the total cross-chain transactions across seven chains. Followed by Polygon, its *Tx_out* is over 8.21 million and the *Tx_in* volume is over 7.17 million, representing 19.23% of the total transaction volume. Table 1 displays the interaction data of cross-chain transactions across seven chains. Following the first two chains, Optimism accounts for 15.95%, BSC for 14.70%, Avalanche for 13.62%, Ethereum for 6.45%, and Fantom for 5.68%. Figure 4 illustrates the time series

of daily transaction volume on seven chains. The significant increase of Arbitrum in April 2024 could be linked to the increase in Figure 3, indicating Stargate transactions are primarily initiated on the Arbitrum blockchain till October 2024. Additionally, we have observed that cross-chain transaction volumes are higher on Layer2 blockchain networks that perform much fewer on-chain consensus, dramatically improving throughput and scalability [75]. These chains attract a large number of DeFi protocols and users with improved scalability and lower fees [26].

4.2 Token Usage

4.2.1 Token Types. Based on the collected data, we identify 8,701 kinds of tokens utilized for cross-chain transactions. This includes 2,999 kinds of tokens on Ethereum, 2,656 on BSC, 1,901 on Polygon, 289 on Avalanche, 618 on Fantom, 289 on Optimism, and 220 on Arbitrum. Additionally, Multichain issues 1,466 anyTokens pegged to the value of ERC20 tokens and WormHole facilitates 4,424 tokens for cross-chain transfers. We consider tokens with the same market information to be the same token type, even if their addresses differ across blockchains. With CoinGecko data serving as the ground truth for identifying tokens of the same type within the cross-chain ecosystem, Table 2 presents information about the top 20 tokens ranked by cross-chain-related trading volume. For completeness, the top 100 tokens are listed in the Appendix, Table 6. Our analysis reveals that on Ethereum, the most popular token for cross-chain transactions is WETH while on Arbitrum, it is SGETH, which is facilitated by the Stargate bridge. On other chains, the most frequently used token for cross-chain transactions is USDC/USDT.

4.2.2 Volume. According to Binance's token rankings, we identify addresses for the top 50 tokens and their corresponding wrapped tokens in the cross-chain ecosystem. These tokens, actively used in 47 million cross-chain transactions, exhibit notable daily volumes on seven chains. For instance, WETH reached $40.82 million, USDC $36.67 million, and USDT $24.69 million. To delve into the significance of cross-chain transactions, we take BSC-USDC as an example, which ranks first in usage. From April 1st to September 1st, 2023, it experienced 248 million transfers, totaling over $195 billion. Among them, 5.9 million were cross-chain transactions, amounting to $2.9 billion or 1.5% of the total value. These findings highlight cross-chain transactions' role in connecting liquidity pools across blockchains, enriching the vibrancy and liquidity of cryptocurrency assets in DeFi.

4.2.3 Active Period. We recognize that notable fluctuations in the frequency of cross-chain token usage over a specific period may signify underlying financial impacts. Consequently, we conduct a study on the frequency of cross-chain tokens and identify distinct active periods for various tokens. Considering each seven days as a time interval, we observe phases marked by sharp increases in trading volume for certain tokens. Interestingly, we find the trading volume for the majority of less mainstream tokens during the initial release period is notably higher than in subsequent periods. After eliminating the initial transfer stage, we obtain more precise results for the active periods. We analyze the daily diversity of bridged tokens and find that the changes in token diversity mirrored the trends in transaction volume. There is a sharp increase and decrease in July 2022, and the period from May to July 2023 sees an increasing variety of tokens used. Overall, the variety of tokens used by users significantly increased from the second half of 2023 compared to previous periods. However, users tend to consistently use USDC for cross-chain operations.

4.3 Users

We next seek to explore the composition and distribution of cross-chain users, based on more than 3.4 million distinct account addresses in our dataset. We first categorize the accounts, before presenting a summary of transactions related to malicious accounts. We then analyze the composition of regular users based on transaction behavior patterns.

Table 2. Top 20 tokens and their related cross-chain transactions.

CAP	Name	Symbol	Trans	CAP	Name	Symbol	Trans
1	USD Coin	USDC	18,689K	11	agEUR	agEUR	232K
2	Tether USD	USDT	8,624K	12	Wrapped BTC	WBTC	201K
3	Stargate Ether Vault	SGETH	7,637K	13	Fantom	WFTM	200K
4	Wrapped Ether	WETH	6,971K	14	Metis Token	Metis	112K
5	UBUSD Token	BUSD	677K	15	BridgeToken:LUNA	(LUNA)	104K
6	Dai Stablecoin	DAI	608K	16	SAND	SAND	104K
7	Magic Internet Money	MIM	480K	17	Wrapped MEMO	wMEMO	97K
8	Wootrade Network	WOO	376K	18	Frax Token	FRAX	84K
9	Wrapped AVAX	WAVAX	316K	19	Genesis	GENESIS	81K
10	Wrapped Matic	WMATIC	274K	20	Governance OHM	gOHM	63K

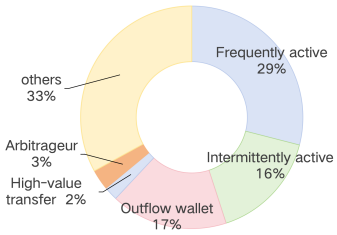


Fig. 6. Distribution of regular users.

4.3.1 Category Definition. Accounts can only be identified by a hash-like address on chains. Thus, it is necessary to classify addresses into different categories to obtain a deeper insight into their characteristics We categorize them by their behavior characteristics in using cross-chain bridges (including frequency, transfer amounts, and transaction roles). Each category is described below:

- **Frequently active user.** Accounts that initiate cross-chain transactions every three days for at least 30 days, or those that engage in cross-chain transactions at least 30% more frequently than other DeFi activities, such as lending and borrowing.
- **Intermittently active user.** Accounts that intermittently exhibit a high frequency of cross-chain transactions, defined as a transaction rate exceeding one standard deviation above the average, with at least seven days between these high-frequency periods.
- **Outflow wallet.** Accounts that primarily transfer funds out via bridges rather than receiving, with a depositor-to-recipient ratio greater than 8:2.
- **High-value transfer.** Accounts that regularly interact with bridges and have pushed cross-chain transactions with a crossed value that exceeds $100K.
- **Arbitrageur.** Accounts where the value of assets received exceeds the assets deposited.

4.3.2 User Composition. We explore the composition of accounts using the above set of categories. First, we extract the transactions of each user and calculate the temporal frequency of cross-chain transactions, thereby identifying frequently/intermittently active users. Second, we calculate the transaction ratio of each account as either the sender or the recipient to identify outflow wallets. Third, we calculate deposited and released asset values for transactions, thereby spotting high-value transfers and arbitrageurs. As a result, we identify that 29.05% of addresses are highly active users, while 16.35% are intermittently active. Additionally, 6% may be arbitrageurs, as the detection results indicate that a portion of their transactions can yield a value on the target chain greater than the deposited value on the source chain. The detailed classification results of regular users based on behavioral characteristics are depicted in Figure 6. We also find that the trend of daily distinct addresses aligns with the cross-chain transaction volume trend.

Answer to RQ1: *Cross-chain transactions has gained a significant popularity, with over 80 million transactions in total. These occupy roughly 5% of transactions across the seven blockchains. The tokens involved in cross-chain transactions show great diversity, with over 8,000 kinds of tokens in total.*

5 Tracking Cross-chain Behaviors

To track the destination and distribution of bridged funds, and to provide insights into the purposes of cross-chain bridge users, we next conduct an analysis of cross-chain fund flows. To this end, we first extract transactions transferring bridged funds on the target chain to build a fund flow graph

for each involved account. Then we collect address labels for account identification. Subsequently, based on graphs and label information, we explore the distribution of cross-chain funds thereby shedding light on user purposes.

5.1 Constructing Fund Flow Graphs

Starting from over 35 million collected cross-in transactions (*Tx_in*s), we construct graphs to capture flow destinations and the related accounts of bridged funds. To standardize the currency, we convert all tokens into their dollar value based on the price of the transaction at the time. The graph construction process is outlined below:

S1. Identify subsequent transactions of *Tx_in*s. Considering each *Tx_in* as an entry, we take its recipient as the sender and search for their transactions involving the bridged token.
S2. Track swapped funds. In most cases, users prefer to use decentralized exchanges to swap tokens to another type, thereby generating subsequent transactions. To enhance the comprehensiveness of the fund flow graphs, we identify a swap transaction with log signatures (*e.g.*, `swap()` in the Uniswap protocol) and parse parameters to extract the recipient, swapped-back token type, and amount. We then take it as a *Tx_in* and return to **S1**, searching for the subsequent transactions.
S3. Terminate recording. The initially bridged tokens, or the tokens identified as swapped in **S2**, may be divided into batches and then be involved in multiple subsequent transactions. To avoid missing or excessive tracking, later transactions will no longer be recorded up until the transaction where the amount of the token transferred equals or exceeds the entry quantity.
S4. Calculate funds value. The market prices of cryptocurrencies are different and undergo daily fluctuations. To compare currency values, we query CoinGecko [8] for token prices and convert all values to dollars.

5.2 Collecting Address Labels

Address labels serve as identification information, assisting blockchain users in recognizing entities like Binance or Coinbase with a hash address. It can also distinguish various labels both on-chain and off-chain, including MEV bots [58], DeFi Whales [32], ENS, as well as pinpoint the wallet type such as MetaMask [17]. Blockchain explorers label accounts to indicate their identity information and these labels are widely recognized in the community, thereby being considered as ground truth to identify malicious users in fund flow graphs.

Within transfer actions in fund flow graphs, we collect 205,464 destination accounts and 6,676 kinds of tokens. To understand the distribution of funds destinations, we are driven to gather label information on involved addresses through *(i)* integrating commercial resources to acquire a fundamental label dataset of on-chain addresses; *(ii)* reorganizing reports to extract exploiter and scammer addresses to compensate for potential omissions; *(iii)* making full use of available label information from CoinGecko and CoinMarketCap [9] to collect addresses related to both Decentralized Exchanges (DEXs) and Centralized Exchanges (CEXs).

Consequently, 73,240,475 labeled addresses are collected, and over 11.56% of the fund destinations are assigned labels. The collected labels can be classified into exchange addresses (*e.g.*, KUCOIN, Coinbase, Binance), malicious accounts, various dApp contracts (*e.g.*, Uniswap, TornadoCash, AAVE), and so on. Out of these, 0.43% are malicious accounts including 3,985 exploiter addresses, 311,233 phishing scammers, and 226 rug pull addresses. and 78% are related to CEXs.

5.3 Distribution of Bridged Funds across Targets

Aiming to explore the target destination and further infer user purposes of interacting with bridges, we first provide an overall impression of the distribution of bridged funds.

DEX funds receipt. A few decentralized cross-chain bridges, such as Multichain and WormHole, issue their own wrapped tokens for thousands of ERC20 tokens, thereby enabling users to withdraw the token of any desired type. That said, most bridges only support a few specific tokens (*e.g.*, USDC, USDT, DAI), which limits what users can directly withdraw on the target chain. As a result, users have to swap their assets into other tokens, and *21% of users prefer to interact with DEXs.* Uniswap [22], ParaSwap [19], AirSwap [6], and Sushi[21] emerge as the top four, accounting for 57.64%, 10.58%, 7.60%, 3.31% of total swapped funds, respectively. This indicates that USDC is the most frequently exchanged target token, and the following is USDT.

CEXs funds receipt. CEXs are another choice for cross-chain users to process their funds and exchange them for cash or other tokens. Users who are less familiar with the process of interacting with smart contracts or directly generating transactions may opt for this method. We identify CEX accounts and find that *3.25% of funds flowed to CEXs*, including Binance, Coinbase, OKEX, ChangeNow *et al.* Among them, the most popular is Binance, accounting for 18.12%, followed by Coinbase where 11.56% of funds flow in.

Funds flow to other dApps. *2.36% of funds flow into other cross-chain bridges* after it has just transferred to the target chain. Besides, some funds are processed to purchase other tokens or NFTs, while *1.17% funds flow to other DeFi dApps* (*i.e.*, Lido, MakerDao, AAVE, *et al.*) to execute liquid staking, lending, *et al.*.

Funds remain in user accounts. Despite various methods to process cross-chain funds, we find that *72.22% funds remain in common accounts* which could either be unprocessed or transferred directly to other common accounts or wallet addresses.

5.4 Inferring the Purposes of Cross-chain Users

From the measurement of destinations and distribution of cross-chain funds, we next make reasonable inferences about user purposes. *(i)* With the insight that most bridged funds were distributed to common users, it can be speculated that users may prefer to spread their resources and assets so as to diminish the risk of a single point of failure or targeted attacks, thereby increasing the resilience of their assets. *(ii)* Observing that funds flow to other dApps after being bridged and, considering the insight that transaction volume is much higher on chains with lower network congestion (see § 4.1.3), it is likely that users desire to lower costs by taking advantage of differences in financial interest rates across different chains. For example, they can borrow tokens from a market deployment on a blockchain featuring a lower interest rate, with the borrowed funds then bridged back to the chain where the loan was opened. *(iii)* Additionally, combined with the 17% percentage of outflow wallets in cross-chain users (see § 4.3.2), we can infer that rewards from DeFi protocols serve as an effective incentive for encouraging users to engage in cross-chain transactions. Those protocols attract capital on different chains so that they can tap into a broader pool of liquidity. *(iv)* Exchanges acting as funds destinations also suggest that users may want to swap native tokens on different blockchains without the necessity of relying on wrapped tokens or CEXs. For example, they can trade ETH on Ethereum for MATIC on Polygon.

5.5 Malicious Accounts with Cross-chain Transactions

It is reported that approximately 3% of the phishing funds are transferred to another chain [49]. Also, some reports [11, 16] demonstrate that a few exploiters tend to process stolen funds via cross-chain transactions. We identify malicious accounts by collecting address labels in Section 5.2.

In this part, malicious accounts are classified into finer-grained categorizations based on the source of their assets including *exploits, rug pulls* [35, 52] and *phishing scams* [49].

Malicious accounts bridge stolen funds. We find that many exploiters aggregate initial funds via cross-chain bridges before executing an attack. After carrying out attacks, stolen funds may also be bridged to other chains. To explore malicious accounts' behavior within cross-chain transactions, we search for their transactions involving stolen funds, which are specified by token address and bridged amount. In our dataset, 576 malicious accounts have completed a total of 1,929 cross-chain transactions via 13 bridges on seven chains, including 775 *Tx_out* and 1,154 *Tx_in*. We also find Ether and stablecoins (*e.g.*, USDC, USDT, and DAI) are most frequently transferred. Specifically, 21,383 Ether and other ERC20 tokens, worth more than $1.12e+9 billion, have been transferred.

Fund flow graphs help track exploited funds. We collect known attacks and scams from DeFiHackLabs [25], Rekt [15], and other publicly available reports. Then we analyze the flow of exploited funds from them, finding that real-world attackers frequently bridge stolen funds. Existing tools do not support tracking cross-chain funds [27, 49], which means either victims or security workers can only depend on manual work when exploited funds are bridged. Our work helps with automated tracking. Given a malicious address, CrossAlert will filter its cross-chain transactions and assist in understanding whether the exploiter obtains initial funds from or transfers exploited funds to other blockchains via bridges. For example, CrossAlert successfully identifies that the ExactlyProtocol exploiter aggregated initial funds using Optimism Bridge, and after the attack, has bridged 4,333 Ether to Ethereum through Across and Optimism Bridge [11]. As well as understanding if attackers bridge exploited funds, fund flow graphs also help track how these funds are handled on the target chain. In this work, results show that after being bridged to another blockchain, over 80% of exploited funds flowed to CEXs, about 2.1% flowed to TornadoCash for money laundering, and the rest remained in common accounts.

Fund flow graphs help block stolen funds and further get recovery. After carrying out attacks, some attackers may transfer funds to other chains, possibly to save subsequent transaction fees or to distribute funds among criminal groups. Our work enables monitoring of attackers' cross-chain operations and immediate identification of relevant accounts on the target chain. This can alert the community and help victims block funds through exchanges, thereby increasing the possibility of recovering stolen assets.

> **Answer to RQ2:** *We discover that 24.42% of cross-chain bridged funds flow to exchanges for a swap on the target chain, and 2.36% flow to other DeFi dApps. As inferred from the fund flow distribution, users likely conduct cross-chain transactions primarily to gain rewards or reduce costs in subsequent DeFi activities. Additionally, we show that fund flow graphs can help track and further block stolen funds.*

6 Detecting Misbehavior

We define misbehaviors in the cross-chain ecosystem as any transaction that involves an unauthorized user triggering gains or losses for a cross-chain bridge. Decentralized cross-chain bridges have always been the target of hackers, resulting in more than $2 billion losses in the last two years [4]. However, apart from these exploits, we conjecture that there are additional misbehaviors that have escaped attention. To bring these to light, we first propose rules that can be inferred from the normal cross-chain process. Based on analyses of transactions, we then extract potential misbehavior patterns.

6.1 Rules for Acceptable Behavior

To identify misbehaviors, we first propose a set of rules that all acceptable transactions should adhere to. Any transactions that fall outside of these rules are assumed to be misbehavior. A cross-chain transaction pair consists of a cross-out transaction (*Tx_out*) and a cross-in transaction (*Tx_in*). On the source chain, a depositor transfers funds to the bridge, and upon receiving accurate funds, the bridge triggers cross-chain-related events to log information including the token contract address, transferred amount, designated target chain, and recipient. While on the target chain, the bridge transfers funds to the recipient after verifying information of *Tx_out*. The following rules can therefore be reasonably inferred from the normal transaction process.

Rule#1 A *Tx_in* must be matched with a corresponding *Tx_out*. This would indicate that a recipient obtaining funds is preceded by the prior deposit of at least funds on the source chain.

Rule#2 Deposited tokens must be supported by the bridge, as listed on their public websites.

Rule#3 Considering factors like service charges and gas fees, the released value to the recipient on the target chain should be no larger than the deposited value on the source chain.

6.2 Detector Methodology and Construction

In order to identify the above misbehaviors from all cross-chain transactions, we propose a comprehensive detection system based on the three pre-defined rules. The detection flow is outlined here. *(i)* According to ***Rule#1*** , we need to first identify *unmatch* transactions. For each *Tx_out*, according to ***Rule#2*** , it is necessary *(ii)* to examine the supportability of a transferred token contract. *(iii)* Once an unsupported token is detected, we must assess the presence of illicit profit. As for matched pairs, according to ***Rule#3*** , *(iv)* our system will create an alert if the deposited amount is less than the released amount, within the same token type. To further detect the existence of arbitrage, *(v)* we compute and compare the real-time finance value of deposited and released tokens, irrespective of their token types.

CrossAlert is composed of four modules: the token inspector, transaction replayer, fund flow builder, and misbehavior detector. We present an overview of CrossAlert in the Figure 7, and describe each component below.

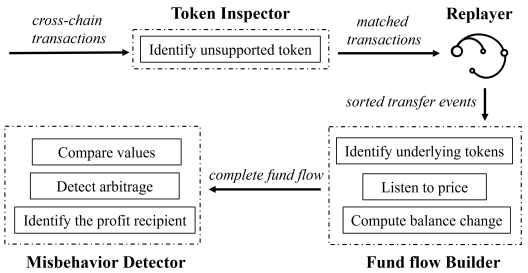


Fig. 7. The overall workflow of CrossAlert.

6.2.1 Token Inspector. The task of this inspector is to determine if the transferred token in a *Tx_out* is unsupported. The process of identifying unsupported tokens primarily relies on officially provided token lists, bridge explorer retrievals, and token deployers for verification.

For each *Tx_out*, this inspector module initially checks if the token is in the list or can be retrieved by the explorer. If not, it proceeds to further identify the token deployer by searching for the first transaction when this token contract was called. Then, our token inspector will consider tokens not deployed from official accounts as risky.

6.2.2 Replayer. The Replayer aims to obtain the complete transaction information from blockchains, and identify all funds transfers within a transaction (including both native token and ERC20 token transfers). Moreover, the Replayer also extracts all transfer-related information for each transfer, including *from_address*, *to_address*, the token contract address, and the transferred amount.

To this end, the Replayer aggregates the external transactions, internal transactions, and triggered events from Geth nodes and blockchain explorers [12]. Subsequently, it filters transfer events with a special signature and parses on-chain bytecode to extract all related information. Ultimately, the Replayer sorts involved transfer events chronologically.

6.2.3 Fund Flow Builder. Based on the sorted transfer events, the Builder constructs a fund flow graph with comprehensive information (*e.g.*, the underlying token for a wrapped one, token decimals, and the real-time price), and computes the change in balance for the accounts involved.

Obtaining token decimals is a straightforward process, yet acquiring real-time market prices for tokens poses some challenges. This is due to the token's daily fluctuations in the market price, coupled with the involvement of multiple token contracts, including wrapped tokens and underlying tokens. In pursuit of this, the Fund Flow Builder *(i)* identifies underlying tokens whose market price can be obtained. For example, imagine that Multichain burns anyWETH in a *Tx_out*, which is a wrapped token with WETH as the underlying. However, in this case, anyWETH may not have a market price, thus necessitating the fund flow builder to identify WETH. To overcome this, the Fund Flow Builder *(ii)* obtains token prices from token information websites, *i.e.*, CoinGecko [8]. Our Fund Flow Builder finally *(iii)* uses the formula 1 to calculate the market value of transferred assets. Here, *Decimals* refers to the number of decimal places for a token. *Amount* represents the data recorded on the blockchain without the decimal point. *Price* signifies the USD value that each of these specific tokens can be exchanged for. Finally, *Value* denotes the dollar value of the token transferred in the transaction.

$$Value = Price * \frac{Amount}{10^{Decimals}} \tag{1}$$

6.2.4 Anomaly Scanner. The task of the anomaly scanner is to generate alerts for unmatched *Tx_in*s and to identify the presence of illicit profits within a complete fund flow graph.

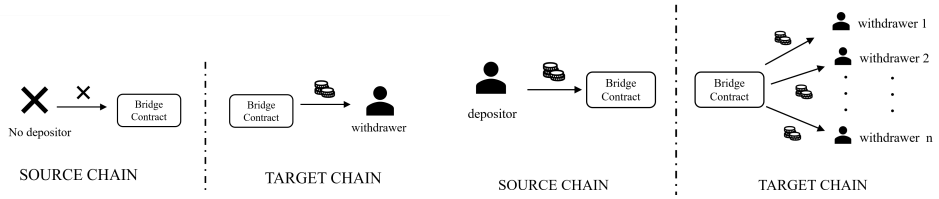
Our anomaly scanner discovers illicit profits by performing comparisons between the token value sent by depositors and the actual value received by the bridge on the source chain. In addition, it also compares deposited token value on the source chain and released token value on the target chain. As for unsupported tokens, the misbehavior detector examines whether they lead to gains for addresses unrelated to the bridge in a cross-chain transaction.

6.3 Analysis and Evaluation

We have deployed CrossAlert to detect real-world misbehaviors by analyzing cross-chain transactions. All previously reported attacks can be accurately pinpointed by CrossAlert. Further, CrossAlert flags more undisclosed misbehaviors, which will be detailed in the following and the summary of detection results is presented in Table 5.

6.3.1 Abnormal Matching. We first look at abnormal matching cases. These transactions violate ***Rule#1*** , since a *Tx_in* cannot be matched with a corresponding *Tx_out*, or multiple *Tx_in*s are matched with the same *Tx_out*.

Explanation. An account directly obtaining funds without any deposits on the source chain is an obvious misbehavior, which is signified by Figure 8a. Another abnormal matching situation is that multiple *Tx_in*s are matched with the same *Tx_out* but the total released funds exceed deposited funds, indicating the existence of illicit profits. Figure 8b is a schematic of this.



(a) A *Tx_in* can not be matched. (b) Multiple *Tx_in*s are matched with the same one.

Fig. 8. Abnormal matching.

Table 3. Abnormal matching. Funds of *transactions cannot be exchanged for US dollars due to low liquidity.

Chain	Transaction Hash	Project	Value	Type
Polygon	0xe3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855	Multichain	$99,999	Unmatch
Ethereum	0xecaaf8b57b6587412242fdc040bd6cc084077a07f4def24b4adae6fbe8254ae3	Multichain	$5,509,227	Unmatch
BSC	0x9e55b7295880dce76aa8af0f3e3f9e36499ae0bdb28088a5924daf29c6132ceb	Multichain	$55,555	Unmatch
Avalanche	0xf0e180c701925623e23c79f76840cbc1069e95454defce5db32e1029104ab572*	Multichain	1e+12B	Unmatch
Avalanche	0x4c12fbface720e0dd5d9cb2f82700e3ed9e32eed0cc8c3b17f094b197bad1385*	Multichain	1e+15	Unmatch
Ethereum	0xa06563fb04698d136c846dfbd762939f2081c4550dbf0ce99a78f78f870183d8	Synapse	$4,999.39	Unmatch

Ground truth. Many bridges provide an official explorer to facilitate retrieving information of a cross-chain transaction pair via a single transaction hash. Transactions that can be retrieved from the explorers are creditable, while those that cannot be retrieved are treated as abnormal ones. As for bridges without explorers, we manually search transactions of the depositor address on the source chain for the accurate *Tx_out*.

Results. In total, we detect 47 unmatched transactions leading to unknown accounts' profits, leading to over $605M loss. By conducting an analysis of unmatched *Tx_in*s, we further discover instances where one *Tx_out* triggered two *Tx_in*s. For example, 0xcb3ab70[5] and 0x0ad7066d[6] are Ethereum *Tx_in*s matched with the same *Tx_out* 0xbce0f56d[7] on BSC, which can be seen in the "srcTxHash" field in the "LogAnySwapIn" log. Ultimately, 47 *Tx_in*s (2 for Celer cBridge, 43 for Multichain, 2 for Synapse) are considered abnormal and some cases are presented in Table 3. Among these, the 43 abnormally matched transactions from Multichain are all due to private key leakage. A hacker observed two transactions to have the same signature and then deduced the private key to Multichain's privileged account in reverse [3]. The other four transactions from Celer cBridge and Synapse were not exposed and we have reported them to the developers. The reply from Celer cBridge suggests the two are due to RPC node errors. However, Synapse developers told us that they cannot figure out the reasons for those unmatched transactions.

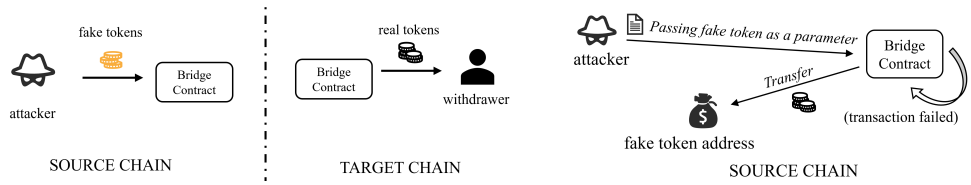
6.3.2 Fake Tokens. We identify cases where bridged tokens violate ***Rule#2***. The token contracts are unsupported by bridges, and the transactions are initiated to exploit smart contracts. **Explanation.** Figure 9 shows two scenarios using unsupported tokens, leading to illicit profits. Tokens not supported by the cross-chain bridge and whose deployer identity that cannot be confirmed are considered risky. However, many transactions involving risky tokens fail the validation of the bridge contract, thus we consider those further causing illicit profits as fake tokens.

Ground truth. Bridges all publish their supporting token lists, which are considered as the ground truth. Note, given the large number of tokens supported by Multichain and WormHole, as well as

[5]https://etherscan.io/tx/0xcb3ab70...

[6]https://etherscan.io/tx/0x0ad7066d...

[7]https://bscscan.com/tx/0xbce0f56d...



(a) Fake token triggers a *Tx_in* thereby releasing funds.

(b) Fake token leads to transaction failure but illicit profits.

Fig. 9. Fake tokens.

Table 4. Fake token information.

Chain	Address	Target Token	Trans
Ethereum	0xb4f89d6a8c113b4232485568e542e646d93cfab1	0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2 (WETH)	362
Avalanche	0xa7f77d387787d279403dc2dddafb6c407005f2d7	0xb31f66aa3c1e785363f0875a1b74e27b85fd66c7 (WAVAX)	111
Avalanche	0x801bd7037dc45b559652ae6e94f6ac7a9a018f1c	0xb31f66aa3c1e785363f0875a1b74e27b85fd66c7 (WAVAX)	76
Ethereum	0xdc9524a8774dc2956bdb8b55fdf91938757f3185	0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2 (WETH)	116

the possibility of contract address updates, there are some tokens not on the lists. Thus, we check the source of funds and examine whether the token deployers are official accounts of the bridge.
Results. In total, 82 fake tokens are identified, where 63 are newly detected by us and 19 are previously reported, leading to over 12.8M ETH loss. We check 8,142 token contracts used by the 13 cross-chain bridges across seven chains. We identify 1,741 tokens that are not on authenticated token lists and have been deployed by unidentified accounts. Yet not all transactions involving these tokens lead to successful exploits. We confirm that 82 addresses, with WETH, PERI, OMT, WBNB, MATIC, and AVAX as underlying tokens, have been passed as tokens to exploit Multichain. These total 1,616 transactions, surpassing a loss of 12.8M Ether, 1M Matic, and 635 WBNB. As for the vulnerability in the Multichain contract, the "*anySwapOutUnderlyingWithPermit*" function doesn't verify whether the "*token*" parameter is indeed a Multichain token contract and it internally invokes the "*permit*" function of the underlying token contract. However, some token contracts (*e.g.*, WETH, PERI, WBNB, MATIC, and AVAX) do not have a permit function but, instead, a fallback function named "deposit" that allows the function's execution to continue normally. This deposits the underlying tokens into the fake token contract and results in an illicit profit.

Regarding publicly available data, 19 token addresses are labeled by blockchain explorers as Multichain exploiters and they are all detected by CrossAlert. After manually checking the transactions involving the other 63 addresses, we confirm them as fake tokens because they are all used to exploit the Multichain vulnerability. We present some fake token examples in Table 4.

6.3.3 Arbitrage. We introduce arbitrage cases that violate ***Rule#3*** . In such cases, a recipient receives an amount no larger than the deposited amount on the source chain, yet can achieve a higher USD value through Center Exchanges (CEXs).
Explanation. As shown in Figure 11, due to the liquidity difference, the same amount of the same token on different blockchains can be exchanged for different USD values.
Results. CrossAlert has performed detection on transactions since April 1st, 2023, identifying 2,005 arbitrage cross-chain transactions, where 18 tokens are involved, leading to over $194K loss. The most common case is where users realize arbitrage through cross-chain transactions involving transferring USDC. To be specific, the market price of USDC is different on different blockchains and may exhibit very minor variations. When the cross-chain amount is large enough, one can

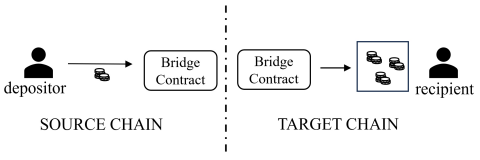


Fig. 10. Extra release.

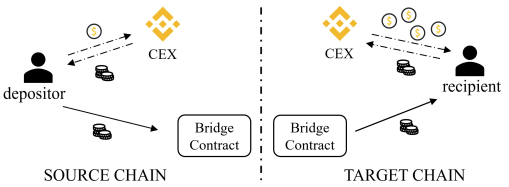


Fig. 11. Arbitrage.

still exchange USDC for more dollars on the target chain even if a portion of the transaction fee is required. CrossAlert also monitored that, around May 20th, 2023, the price of *Woo Network Token* on BSC was \$0.23, whereas, on the Polygon, it was four times higher, at \$0.88. This stimulated at least 366 cross-chain transactions from Polygon to BSC (*e.g.*, *Tx_out*: 0xe6e676[8], corresponding *Tx_in*: 0x583abe[9]; *Tx_out*: 0x69ca79[10], corresponding *Tx_in*: 0xe0ba87[11]), resulting in over \$164K profit. Many traders performed cross-chain transactions in this period for arbitrage opportunities. They transferred assets from BSC to Polygon and then swapped *WOO NETWORK* token to stablecoins.

6.3.4 Extra Release. We next look at cases where transactions violate ***Rule#3*** . This covers cases where, regardless of token price, the amount transferred to recipients on the target chain is larger than the amount deposited on the source chain. This clearly indicates a type of misbehavior.

Explanation. The attacker may have legitimate keys to sign messages, or there may be a bug in the signing process of the exploited bridge that has been abused to sign a crafted message. The ultimate result is that the bridge contract signs and authorizes transactions for transferring a larger amount than the deposited amount.

Result. In total, 119 transactions are detected, leading to \$10.1M loss. CrossAlert detects the exploit on PolyNetwork on July 2nd, 2023 in real time. From this, we find that 57 different assets across 10 different blockchains were involved in this attack and the exploiter exchanged a total of 5,196 ETH (approximately \$10.1 million) on Ethereum. Taking one *Tx_out* 0x249aaf and its matched *Tx_in* 0xef767b as an example, the attacker deposited $1 * 10^{-12}$ OOE Token on Ethereum but withdrew over 786 million OOE on BSC.

Transactions related to the attack on PolyNetwork are within the scope of the seven chains and they are all detected by CrossAlert. However, attacks outside of the chains (*e.g.*, attacks on WormHole on Solona, and PolyNetwork with Curve as the source chain) cannot be detected. We confirm that the attack on PolyNetwork was carried out by generating signatures using the project's wallets. The attack on WormHole is due to an attacker exploiting the vulnerability in cross-chain protocols to bypass verification.

6.3.5 Summary. Table 5 summarizes the detection results. CrossAlert successfully identifies 100% of the transactions associated with previously reported attacks on the 13 bridges. Further, CrossAlert have flagged hundreds of undisclosed misbehaviors[12]. Specifically, CrossAlert detects 82 fake token addresses that are involved in Multichain exploits. Among these, all addresses that have been labeled by blockchain explorers are detected and, despite the labeled (reported) 19 tokens, we additionally identify 63 undisclosed token addresses. Moreover, CrossAlert detects 94 *Tx_in*s that are abnormal in the matching process. Additionally, we detect over 2K arbitrage transactions exploiting value

[8]https://bscscan.com/tx/0xe6e676...
[9]https://polygonscan.com/tx/0x583abe...
[10]https://optimistic.etherscan.io/tx/0x69ca79...
[11]https://polygonscan.com/tx/0xe0ba87...
[12]We are in the process of reporting issues to the corresponding projects.

Table 5. A summary of detected misbehaviors.

Type	Trans	Least loss	Previous Reported	Newly Discovered
Fake tokens	1,616	12.8M ETH	19	63
Abnormal matching	94	$605M	0	94
Arbitrage	2,005	$194K	0	2,005
Extra release	119	$ 10.1M	119	0
Sum	3,863	$615.5M and 12.8M ETH		

differentials across chains for arbitrage. 18.2% of these yield an impressive profit at most 3x more than the deposited funds.

Summarizing root causes for all exposed misbehaviors, we find that most *abnormal matching* and *extra release* are carried out due to private key leakages. Other reasons may include protocol vulnerabilities or DNS vulnerabilities [78]. Attacks using *fake tokens* primarily exploit contract vulnerabilities. Detailed analysis of other cross-chain bridge attacks beyond our study scope has been demonstrated by Lee [55].

For mitigation, safeguarding the private keys is of utmost importance, and it is also crucial to ensure robust access control. Additionally, using wrapped tokens can introduce risks, so comprehensive implementation tests should be employed. Maintaining token pair maps can efficiently prevent attackers from bypassing verification without access control.

Answer to RQ3: *We have summarized four observable misbehavior patterns within the cross-chain ecosystem and built CrossAlert to detect real-world misbehaviors. All transactions involved in previously reported attacks on the 13 bridges are successfully detected. We additionally identify 82 fake tokens, exploiting more than 12.8M ETH, and detect 94 abnormally matched transactions. CrossAlert also detects 366 significant cross-chain arbitrage transactions, exploiting the value differentials of WOO tokens across various chains for over $164K profit.*

7 Discussion

7.1 Implications

Our observations hold significance for multiple stakeholders in the community. First, our work helps to address the challenges of cross-chain data analysis caused by non-uniform function interfaces of different bridge contracts, so as to further assist regulatory efforts, which also implies an appeal to standard compliance. Second, our data is significant for supporting efficient transaction data retrieval in the decentralized cross-chain ecosystem, which can further allow the community to build comprehensive datasets and track funds automatically. Third, we detected security issues within the cross-chain ecosystem and explored their root causes. It should be noted that enhanced contract robustness and privileged account security measures, such as multi-signature schemes, are essential to mitigate the risk of exploitation. Measurements in our work will foster collaboration among developers, regulators, and researchers, driving the development of best practices and strengthening the overall security and efficiency of the blockchain ecosystem.

7.2 Limitations

Our work is subject to several limitations. First, we rely on heuristics and manual validation due to the limited availability of ground-truth data. Notably, when collecting attacker and scammer addresses, we depend on labels from blockchain explorers and media reports, making it difficult to ensure the collection of all malicious accounts. However, our investigation indicates the proportion of malicious addresses using cross-chain bridges is inherently low, thus, a minimal number of

missed addresses does not significantly impact our research. Second, our research is focused solely on transactions on seven chains, which brings a limitation to our research scope. This is because, at the beginning of our work, data for only seven chains was available from blockchain explorers and Chainbase. Subsequent efforts can augment data from additional chains. Third, despite our extensive efforts to search for relevant reports or contact project teams, we cannot systematically trace the root causes of detected in-the-wild unmatched *Tx_in*s.

7.3 CrossAlert Operation Overheads

We will make the code for collecting and matching cross-chain transactions public. However, the code for the detector will not be disclosed, as the detector's replayer and fund flow builder are built upon commercial software partnerships and confidentiality agreements. Unfortunately, since CoinGecko no longer supports free retrieval for token prices, it is difficult for us to offer the service for free. If anybody wishes to run this methodology themselves, it entails two main overheads. First, using CoinGecko to obtain token prices comes with a financial cost of at least $129 per month. As an alternative, it is possible to use Uniswap to calculate token prices (which is free). However, this approach has lower coverage for cross-chain tokens across the seven blockchains, and there may be scam tokens on Uniswap [74]. Second, simulating transactions is necessary for constructing the Replayer and the Fund Flow Builder, yet both local simulation and third-party services involve overheads. If locally simulating a historical transaction, it is typically necessary to maintain a full blockchain node, storing all historical data, including blocks and transactions. This incurs significant storage costs. It is usually easier to employ third-party simulators to organize fund flows and calculate balance changes, but this will introduce a minimum delay of 5 seconds per transaction and increase the network burden. Naturally, somebody wishing to deploy CrossAlert can select the setup that best reflects their preferences.

8 Related Work

8.1 Cross-chain Ecosystem

Zamyatin *et al.* [76] study communication between blockchains, necessary for the communicator component of a bridge. Instances of cross-chain communication that involve proving the state of one ledger to another are explored in related works like [29] and [30]. Other systems, such as Cosmos [54] and Polkadot [73], aim to support cross-chain communications and transfers as core functionality. Han *et al.* [48] generate a comprehensive set of criteria on security, privacy, and other performance, providing a review of existing cross-chain technologies. Deng *et al.* [41] focus on analyzing side-chain and hash-lock technologies. Besides, some researchers provide a retrospective analysis of attacks against the cross-chain system and also summarize defense approaches [43, 55]. Yet, there is no work on investigating the cross-chain ecosystem from a comprehensive perspective. This paper is the first to conduct an empirical study by analyzing millions of transactions from multiple decentralized cross-chain bridges.

8.2 Transaction-based Analysis of Blockchains

Some existing research studies have undertaken transaction-based analyses to delve into blockchain systems [37, 56, 80]. Notably, investigations into Bitcoin [42, 45, 63, 65, 67, 79] have focused on aspects such as de-anonymization and the utilization of graph-based methodologies for detecting money laundering. Also, some researchers have explored EOSIO and Counterfeit Cryptocurrency through transaction-based inquiries [50, 53]. In a separate study, Gao *et al.* [47] analyze over 190K ERC-20 tokens to investigate the presence of counterfeit cryptocurrencies on Ethereum and measure their impact. Furthermore, Wang *et al.* [71] conduct the first empirical study to quantify the risk

associated with the unlimited approval of ERC20 tokens on Ethereum. However, to the best of our knowledge, there remains a lack of research that delves deeply into the analysis of cross-chain transactions and their characteristics.

8.3 Vulnerability/Attack Detection of Blockchains

The blockchain ecosystem has encountered numerous security challenges since its inception, prompting various research initiatives aimed at quantifying and gaining a deeper understanding of these issues. SADPonzi [38] proposes the detection of Ponzi scams in smart contracts by extracting semantic information through symbolic execution and comparing it to summarized Ponzi scheme patterns. He *et al.* [50] characterize the code reuse practice in the Ethereum smart contract ecosystem. For the detection of scam tokens, Xia *et al.* [74] suggest training a classifier by extracting time-series, transaction, investor, and Uniswap-specific features. In the realm of detecting phishing scams through transactions on Ethereum, He *et al.* [49] conduct the first empirical study, successfully detecting and reporting 26,333 TxPhish websites and 3,486 phishing accounts. In the field of cross-chain, several studies focus on the security of atomic cross-chain swap protocols [51, 70], with only Xscope[77] proposed to find security violations in cross-chain bridges.

9 Conclusion

This paper presents the first large-scale measurement study of the cross-chain ecosystem. We have built a comprehensive and accurate dataset encompassing over 80 million cross-chain transactions from 13 decentralized bridges on seven blockchains and completed matching transaction pairs. We then unveiled the prevalence of this ecosystem by measuring transaction volume, token usage, as well as user behaviors. Further, we constructed funds flow graphs to explore user purposes and track crossed funds of malicious accounts. Additionally, we have discussed four types of observable misbehaviors and constructed CrossAlert for detection. Our observations, dataset and tools can positively contribute to the community and boost a series of research studies on cross-chain.

Acknowledgments

We sincerely thank our shepherd Prof. Andrea Marin (Università Ca' Foscari di Venezia) and all the anonymous reviewers for their valuable suggestions and comments to improve this paper. This work was supported in part by the National Key R&D Program of China (2021YFB2701000), the Key R&D Program of Hubei Province (2023BAB017, 2023BAB079), the National NSF of China (grants No.62072046, 62302181), the Knowledge Innovation Program of Wuhan-Basic Research (2022010801010083), the Xiaomi Young Talents Program, and the HUSTCSE-FiberHome Joint Research Center for Network Security.

References

[1] [n. d.]. *Chainbase.* https://chainbase.com/ September 30, 2023.
[2] [n. d.]. *Geth.* https://geth.ethereum.org/ September 30, 2023.
[3] 2021. *Anyswap Multichain Router V3 Exploit Statement.* Retrieved July,27, 2024 from https://medium.com/multichainorg/anyswap-multichain-router-v3-exploit-statement-6833f1b7e6fb
[4] 2022. *Vulnerabilities in Cross-chain Bridge Protocols Emerge as Top Security Risk.* Retrieved January 5, 2023 from https://www.chainalysis.com/blog/cross-chain-bridge-hacks-2022/
[5] 2023. *Across protocol official website.* Retrieved December 14, 2023 from https://across.to/
[6] 2023. *AirSwap.* https://www.airswap.io/
[7] 2023. *Arbitrum Bridge.* Retrieved September 30, 2023 from https://bridge.arbitrum.io/
[8] 2023. *CoinGecko.* Retrieved December 14, 2023 from https://www.coingecko.com//
[9] 2023. *CoinMarketCap.* Retrieved Dec 30, 2023 from https://coinmarketcap.com/api/
[10] 2023. *Cross-chain Bridges Dashboard.* Retrieved September 30, 2023 from https://defillama.com/bridges

[11] 2023. *Defi 'Exactly Protocol' Hack Analysis.* Retrieved December 14, 2023 from https://www.immunebytes.com/blog/defi-exactly-protocol-hack-analysis/

[12] 2023. *Etherscan.* Retrieved April 1, 2023 from https://etherscan.io

[13] 2023. *Five of top 10 blockchain exploits in 2022 are related to cross-chain.* Retrieved September 30, 2023 from https://www.cybavo.com/blog/year-in-review-top-defi-hacks-2022/

[14] 2023. *Hop Exchange.* Retrieved September 30, 2023 from https://app.hop.exchange/#/send?sourceNetwork=ethereum

[15] 2023. *https://de.fi/rekt-database.* Retrieved Dec 30, 2023 from https://de.fi/rekt-database/

[16] 2023. *Magnate Finance Scammer has bridged the majority of the profits out.* Retrieved December 14, 2023 from https://twitter.com/MetaSleuth/status/1695005203920253414

[17] 2023. *MetaMask.* Retrieved Mar 1, 2023 from https://metamask.io

[18] 2023. *Multichain Explorer.* Retrieved September 30, 2023 from https://scan.multichain.org/#/

[19] 2023. *ParaSwap: Best Prices in DeFi for Traders & dApps.* https://www.paraswap.xyz/

[20] 2023. *Stargate finance official website.* Retrieved December 14, 2023 from https://stargate.finance/overview

[21] 2023. *Sushi.* https://www.sushi.com/

[22] 2023. *Uniswap Protocol.* Retrieved Dec 10, 2023 from https://uniswap.org/

[23] 2023. *What is Ethereum.* Retrieved February 26, 2023 from https://ethereum.org/en/what-is-ethereum

[24] 2024. *DeFi TVL Returns To $100 Billion.* Retrieved June,13, 2024 from https://www.theblock.co/post/280657/defi-tvl-breaks-100-billion-for-first-time-since-may-2022

[25] 2024. *DeFiHackLabs's Substack.* Retrieved August 3, 2024 from https://defihacklabs.substack.com/

[26] 2024. *A Look at Polygon (MATIC): Why Layer 2 Blockchains are the Future.* Retrieved June,13, 2024 from https://blog.bake.io/layer-2-blockchains-polygon/

[27] 2024. *MetaSleuth: Crypto Tracking and Investigation Platform.* Retrieved July 14, 2024 from https://metasleuth.io/

[28] 2024. *What is Solana?* Retrieved June,13, 2024 from https://solana.com/docs/intro/overview

[29] Ermyas Abebe, Dushyant Behl, Chander Govindarajan, Yining Hu, Dileban Karunamoorthy, Petr Novotny, Vinayaka Pandit, Venkatraman Ramakrishna, and Christian Vecchiola. 2019. Enabling enterprise blockchain interoperability with trusted data transfer (industry track). In *Proceedings of the 20th international middleware conference industrial track*. 29–35.

[30] Ermyas Abebe, Yining Hu, Allison Irvin, Dileban Karunamoorthy, Vinayaka Pandit, Venkatraman Ramakrishna, and Jiangshan Yu. 2021. Verifiable observation of permissioned ledgers. In *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 1–9.

[31] Fadi Barbàra and Claudio Schifanella. 2022. BxTB: cross-chain exchanges of bitcoins for all Bitcoin wrapped tokens. In *2022 Fourth International Conference on Blockchain Computing and Applications (BCCA)*. IEEE, 143–150.

[32] Tom Barbereau, Reilly Smethurst, Orestis Papageorgiou, Alexander Rieger, and Gilbert Fridgen. 2022. Defi, not so decentralized: The measured distribution of voting rights. (2022).

[33] Rafael Belchior, André Vasconcelos, Sérgio Guerreiro, and Miguel Correia. 2021. A survey on blockchain interoperability: Past, present, and future trends. *ACM Computing Surveys (CSUR)* 54, 8 (2021), 1–41.

[34] Vitalik Buterin. 2016. Chain interoperability. *R3 research paper* 9 (2016), 1–25.

[35] Federico Cernera, Massimo La Morgia, Alessandro Mei, and Francesco Sassi. 2023. Token Spammers, Rug Pulls, and Sniper Bots: An Analysis of the Ecosystem of Tokens in Ethereum and in the Binance Smart Chain (BNB). In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 3349–3366. https://www.usenix.org/conference/usenixsecurity23/presentation/cernera

[36] Tao-Hung Chang and Davor Svetinovic. 2016. Data analysis of digital currency networks: Namecoin case study. In *2016 21st International Conference on Engineering of Complex Computer Systems (ICECCS)*. IEEE, 122–125.

[37] Ting Chen, Zihao Li, Yuxiao Zhu, Jiachi Chen, Xiapu Luo, John Chi-Shing Lui, Xiaodong Lin, and Xiaosong Zhang. 2020. Understanding ethereum via graph analysis. *ACM Transactions on Internet Technology (TOIT)* 20, 2 (2020), 1–32.

[38] 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. *Proc. ACM Meas. Anal. Comput. Syst.* 5, 2, Article 26 (June 2021), 30 pages. https://doi.org/10.1145/3460093

[39] Raymond Cheng, Fan Zhang, Jernej Kos, Warren He, Nicholas Hynes, Noah Johnson, Ari Juels, Andrew Miller, and Dawn Song. 2019. Ekiden: A platform for confidentiality-preserving, trustworthy, and performant smart contracts. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 185–200.

[40] Pau Cuesta Arcos. 2023. *Analysis of bridge-solutions for public blockchains.* Master's thesis. Universitat Politècnica de Catalunya.

[41] Liping Deng, Huan Chen, Jing Zeng, and Liang-Jie Zhang. 2018. Research on cross-chain technology based on sidechain and hash-locking. In *Edge Computing–EDGE 2018: Second International Conference, Held as Part of the Services Conference Federation, SCF 2018, Seattle, WA, USA, June 25-30, 2018, Proceedings 2*. Springer, 144–151.

[42] Damiano Di Francesco Maesa, Andrea Marino, and Laura Ricci. 2017. An analysis of the bitcoin users graph: inferring unusual behaviours. In *Complex Networks & Their Applications V: Proceedings of the 5th International Workshop on Complex Networks and their Applications (COMPLEX NETWORKS 2016)*. Springer, 749–760.

[43] Li Duan, Yangyang Sun, Wei Ni, Weiping Ding, Jiqiang Liu, and Wei Wang. 2023. Attacks Against Cross-Chain Systems and Defense Approaches: A Contemporary Survey. *IEEE/CAA Journal of Automatica Sinica* 10, 8 (2023), 1647–1667.

[44] Jacob Eberhardt and Stefan Tai. 2018. Zokrates-scalable privacy-preserving off-chain computations. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 1084–1091.

[45] Michael Fleder, Michael S Kester, and Sudeep Pillai. 2015. Bitcoin transaction graph analysis. *arXiv preprint arXiv:1502.01657* (2015).

[46] Ankit Gangwal, Haripriya Ravali Gangavalli, and Apoorva Thirupathi. 2023. A survey of Layer-two blockchain protocols. *Journal of Network and Computer Applications* 209 (2023), 103539.

[47] Bingyu Gao, Haoyu Wang, Pengcheng Xia, Siwei Wu, Yajin Zhou, Xiapu Luo, and Gareth Tyson. 2020. Tracking Counterfeit Cryptocurrency End-to-end. *Proc. ACM Meas. Anal. Comput. Syst.* 4, 3, Article 50 (Nov. 2020), 28 pages. https://doi.org/10.1145/3428335

[48] Panpan Han, Zheng Yan, Wenxiu Ding, Shufan Fei, and Zhiguo Wan. 2023. A survey on cross-chain technologies. *Distributed Ledger Technologies: Research and Practice* 2, 2 (2023), 1–30.

[49] 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* (Copenhagen, Denmark) *(CCS '23)*. Association for Computing Machinery, New York, NY, USA, 120–134. https://doi.org/10.1145/3576915.3623210

[50] Ningyu He, Ruiyi Zhang, Haoyu Wang, Lei Wu, Xiapu Luo, Yao Guo, Ting Yu, and Xuxian Jiang. 2021. EOSAFE: Security Analysis of EOSIO Smart Contracts. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 1271–1288. https://www.usenix.org/conference/usenixsecurity21/presentation/he-ningyu

[51] Maurice Herlihy. 2018. Atomic cross-chain swaps. In *Proceedings of the 2018 ACM symposium on principles of distributed computing*. 245–254.

[52] Jintao Huang, Ningyu He, Kai Ma, Jiang Xiao, and Haoyu Wang. 2023. A Deep Dive into NFT Rug Pulls. arXiv:2305.06108 [cs.CR]

[53] Yuheng Huang, Haoyu Wang, Lei Wu, Gareth Tyson, Xiapu Luo, Run Zhang, Xuanzhe Liu, Gang Huang, and Xuxian Jiang. 2020. Understanding (Mis)Behavior on the EOSIO Blockchain. *Proc. ACM Meas. Anal. Comput. Syst.* 4, 2, Article 37 (June 2020), 28 pages. https://doi.org/10.1145/3392155

[54] Jae Kwon and Ethan Buchman. 2019. Cosmos whitepaper: A network of distributed ledgers. *White Paper. Available online: https://cosmos. network/resources/whitepaper (accessed on 13 July 2021)* 8 (2019).

[55] Sung-Shine Lee, Alexandr Murashkin, Martin Derka, and Jan Gorzny. 2023. SoK: Not Quite Water Under the Bridge: Review of Cross-Chain Bridge Hacks. In *2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. 1–14. https://doi.org/10.1109/ICBC56567.2023.10174993

[56] Xi Tong Lee, Arijit Khan, Sourav Sen Gupta, Yu Hann Ong, and Xuan Liu. 2020. Measurements, analyses, and insights on the entire ethereum blockchain network. In *Proceedings of The Web Conference 2020*. 155–166.

[57] Ankur Lohachab, Saurabh Garg, Byeong Kang, Muhammad Bilal Amin, Junmin Lee, Shiping Chen, and Xiwei Xu. 2021. Towards interconnected blockchains: a comprehensive review of the role of interoperability among disparate blockchains. *ACM Computing Surveys (CSUR)* 54, 7 (2021), 1–39.

[58] Bruno Mazorra, Michael Reynolds, and Vanesa Daza. 2022. Price of MEV: towards a game theoretical approach to MEV. In *Proceedings of the 2022 ACM CCS Workshop on Decentralized Finance and Security*. 15–22.

[59] Andrew Miller, Iddo Bentov, Surya Bakshi, Ranjit Kumaresan, and Patrick McCorry. 2019. Sprites and state channels: Payment networks that go faster than lightning. In *International conference on financial cryptography and data security*. Springer, 508–526.

[60] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized business review* (2008).

[61] Lightning Network. 2018. Lightning network: scalable, instant Bitcoin/blockchain transactions.

[62] Wei Ou, Shiying Huang, Jingjing Zheng, Qionglu Zhang, Guang Zeng, and Wenbao Han. 2022. An overview on cross-chain: Mechanism, platforms, challenges and advances. *Computer Networks* 218 (2022), 109378.

[63] Silivanxay Phetsouvanh, Frédérique Oggier, and Anwitaman Datta. 2018. Egret: Extortion graph exploration techniques in the bitcoin network. In *2018 IEEE International conference on data mining workshops (ICDMW)*. IEEE, 244–251.

[64] Tomas Rafaj, Lukas Mastilak, Kristian Kostal, and Ivan Kotuliak. 2023. DeFi Gaming Platform Using the Layer 2 Benefits. In *2023 33rd Conference of Open Innovations Association (FRUCT)*. IEEE, 236–242.

[65] Fergal Reid and Martin Harrigan. 2013. *An analysis of anonymity in the bitcoin system*. Springer.

[66] Peter Robinson. 2021. Survey of crosschain communications protocols. *Computer Networks* 200 (2021), 108488.

[67] Dorit Ron and Adi Shamir. 2013. Quantitative analysis of the full bitcoin transaction graph. In *Financial Cryptography and Data Security: 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers 17*. Springer, 6–24.

[68] Cosimo Sguanci, Roberto Spatafora, and Andrea Mario Vergani. 2021. Layer 2 blockchain scaling: A survey. *arXiv preprint arXiv:2107.10881* (2021).

[69] Josh Stark. 2018. Making sense of ethereum's layer 2 scaling solutions: State channels, plasma, and truebit. *Medium. com* (2018).

[70] Itay Tsabary, Matan Yechieli, Alex Manuskin, and Ittay Eyal. 2021. MAD-HTLC: Because HTLC is Crazy-Cheap to Attack. In *2021 IEEE Symposium on Security and Privacy (SP)*. 1230–1248. https://doi.org/10.1109/SP40001.2021.00080

[71] Dabao Wang, Hang Feng, Siwei Wu, Yajin Zhou, Lei Wu, and Xingliang Yuan. 2022. Penny wise and pound foolish: quantifying the risk of unlimited approval of ERC20 tokens on ethereum. In *Proceedings of the 25th International Symposium on Research in Attacks, Intrusions and Defenses*. 99–114.

[72] Hui Wang, Yuanyuan Cen, and Xuefeng Li. 2017. Blockchain router: A cross-chain communication protocol. In *Proceedings of the 6th international conference on informatics, environment, energy and applications*. 94–97.

[73] Gavin Wood. 2016. Polkadot: Vision for a heterogeneous multi-chain framework. *White paper* 21, 2327 (2016), 4662.

[74] 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.

[75] Zihuan Xu and Lei Chen. 2022. L2chain: Towards High-Performance, Confidential and Secure Layer-2 Blockchain Solution for Decentralized Applications. *Proc. VLDB Endow.* 16, 4 (dec 2022), 986–999. https://doi.org/10.14778/3574245.3574278

[76] Alexei Zamyatin, Mustafa Al-Bassam, Dionysis Zindros, Eleftherios Kokoris-Kogias, Pedro Moreno-Sanchez, Aggelos Kiayias, and William J Knottenbelt. 2021. Sok: Communication across distributed ledgers. In *Financial Cryptography and Data Security: 25th International Conference, FC 2021, Virtual Event, March 1–5, 2021, Revised Selected Papers, Part II 25*. Springer, 3–36.

[77] Jiashuo Zhang, Jianbo Gao, Yue Li, Ziming Chen, Zhi Guan, and Zhong Chen. 2022. Xscope: Hunting for Cross-Chain Bridge Attacks. arXiv:2208.07119 [cs.SE]

[78] Mengya Zhang, Xiaokuan Zhang, Josh Barbee, Yinqian Zhang, and Zhiqiang Lin. 2023. SoK: Security of Cross-chain Bridges: Attack Surfaces, Defenses, and Open Problems. *arXiv preprint arXiv:2312.12573* (2023).

[79] Chen Zhao and Yong Guan. 2015. A graph-based investigation of bitcoin transactions. In *Advances in Digital Forensics XI: 11th IFIP WG 11.9 International Conference, Orlando, FL, USA, January 26-28, 2015, Revised Selected Papers 11*. Springer, 79–95.

[80] Lin Zhao, Sourav Sen Gupta, Arijit Khan, and Robby Luo. 2021. Temporal analysis of the entire ethereum blockchain network. In *Proceedings of the Web Conference 2021*. 2258–2269.

Appendix for the work

Table 6. Top 100 Tokens in Cross-chain Ecosystem.

CAP	Name	Symbol	Trans	CAP	Name	Symbol	Trans
1	USD Coin	USDC	18689179	51	Lyra Token	LYRA	11603
2	Tether USD	USDT	8623831	52	LIQR	LIQR	11014
3	Stargate Ether Vault	SGETH	7636888	53	GMX	GMX	10104
4	Wrapped Ether	WETH	6970434	54	MCDEX Token	MCB	9956
5	UBUSD Token	BUSD	676410	55	Kromatika	KROM	9897
6	Dai Stablecoin	DAI	607997	56	MIMO Parallel Governance Token	MIMO	9870
7	Magic Internet Money	MIM	480200	57	Palette Token	PLT	9657
8	Wootrade Network	WOO	375333	58	Balancer	BAL	9577
9	Wrapped AVAX	WAVAX	315248	59	MagicCraft	MCRT	9458
10	Wrapped Matic	WMATIC	273026	60	Synapse Network	SNP	9425
11	agEUR	agEUR	231792	61	Thales DAO Token	THALES	9279
12	Wrapped BTC	WBTC	200805	62	Forta	FORT	9089
13	Fantom	WFTM	199033	63	Decentral Games	DG	8962
14	Metis Token	Metis	112152	64	Orion Money Token	ORION	8927
15	BridgeToken:LUNA	(LUNA)	103326	65	Curve DAO Token	CRV	8360
16	SAND	SAND	103040	66	LUSD Stablecoin	LUSD	8198
17	Wrapped MEMO	wMEMO	96223	67	XANA	XETA	7942
18	Frax Token	FRAX	83491	68	Aave Token	AAVE	7863
19	Genesis	GENESIS	80779	69	DUST Protocol	DUST	7790
20	Governance OHM	gOHM	63072	70	HyperJump	JUMP	7562
21	JEWEL	JEWEL	50389	71	Domi	DOMI	7349
22	Synapse	SYN	49761	72	Pendle	PENDLE	6883
23	Spell Token	SPELL	47857	73	GOVI	GOVI	6653
24	Inverse ETH Volatility Index	iETHV	37741	74	THE TRUTH	UFO	6601
25	Mai Stablecoin	MAI	35834	75	Telcoin	TEL	6599
26	Ripae	PAE	33923	76	DFX Token	DFX	6564
27	Meta Apes Peel	PEEL	32540	77	Lido DAO Token	LDO	6550
28	beefy.finance	BIFI	31261	78	CherrySwapToken	CHE	6535
29	Treasure Under Sea	TUS	27268	79	Wing Token	WING	6320
30	DeRace Token	DERC	26171	80	Dogecoin	DOGE	6197
31	DeRace Token	DERC	26171	81	Kalmar Token	KALM	5835
32	Wrapped SOL	SOL	24750	82	GAMEE	GMEE	5816
33	StackOS	STACK	24321	83	Arable Protocol	ACRE	5791
34	TOMB	TOMB	19269	84	NFT Worlds	WRLD	5686
35	CryptoBlades Skill Token	SKILL	17290	85	PANCAKE GAMES	GCAKE	5375
36	DFX Token	DFX	16253	86	GemGuardian	GEMG	5119
37	Frax Share	FXS	15468	87	Mantle	MNT	4943
38	Hundred Finance	HND	15179	88	Aavegotchi GHST Token	GHST	4885
39	Kyber Network Crystal	KNC	15067	89	Rally	RLY	4835
40	Qi Dao	QI	14790	90	pTokens TLOS	TLOS	4424
41	STEP.APP	FITFI	14334	91	MetaShooter	MHUNT	4262
42	ChainLink Token	LINK	14142	92	Graph Token	GRT	4062
43	Mad Meerkat Finance	MMF	13540	93	MCHCoin	MCHC	3951
44	Karmaverse Zombie Serum	Serum	13484	94	Annex	ANN	3740
45	DEUS	DEUS	12738	95	Livepeer Token	LPT	3594
46	Poly-Peg MDX	HMDX	12182	96	ZEBEC	ZBC	3399
47	Oath Token	OATH	12150	97	BSC Conflux	bCFX	3171
48	Tarot	TAROT	12011	98	Wombat Token	WOM	3044
49	O3 Swap Token	O3	11957	99	Star Atlas	ATLAS	2928
50	Mobius Token	MOT	11652	100	Yup	YUP	2736

Received August 2024; revised September 2024; accepted October 2024

Table 7. Key Fields For Cross-chain Transactions.

Field	Explanation
from_chain	the source chain network
from_address	the address depositing assets on the source chain
from_token	the token locked by the bridge on the source chain
from_amount	locked raw amount by the bridge on the source chain
to_chain	the target chain network
to_address	the address withdrawing assets on the target chain
to_token	the token issued by the bridge on the target chain
to_amount	issued raw amount by the bridge on the target chain
match_tag	designed for precisely matching corresponding transactions pairs
direction	specific for distinguishing if it is a *Tx_out* or *Tx_in*