Anti-Sybil Project

To Build an Intelligent Sybil Discovery LEGO

Trusta Labs

Agenda

1. Introduction
    1. Team
    2. Deliverables (github & demo video)
2. Our Works
    1. Data Preparation
    2. Topic 1: Bulk Transfers & Donations
    3. Topic 2: Sequential Behavior Pattern Mining
    4. Topic 3 : Asset-Transfer Graph Mining
    5. Topic 4: Grant Fraud
3. Summary, Suggestion and Future Works

Team and Deliverables

Team

We are a team of web3 data scientists aiming at preventing sybil attacks. Before the Gitcoin Hackthon, we have already done a lot of preparations and related works, such as the Sybil analysis on HOP and Gnosis Safe airdrop. We position our work as the algorithmic detection LEGO in a Sybil resistance system although it means high development and maintenance cost.

Deliverables

The ppt contains our thinking, logic, methodology, algorithms and reasoning on a bunch of cases.

We have also uploaded the code to collect data, conduct feature engineering, compute risk score and Sybil clusters, and generate visualizations.

We keep on building an anti-sybil system. Look forward to collaborating with data scientists from Gitcoin community in Sybil hunting.

Agenda

Data Preparation

Collected Data

- Gitcoin-Hackthon: GR15 donation detailed info GR_15_DATA
- Alchemy: Ethereum and Polygon transfers related to GR15 contributors
- zkSync: L2 transfers related to GR15 contributors
- Chainbase: token prices in USD
- DUNE: GR15 address tags from "GR15-bignode-name"

Table	Description	Size
grants_applications	grants info and description	808
contributions_dataset	Contributions record	475046
grants	List of grants	1503
experiment_on_chain_data	On-chain info about a small number of contributors	35
experiment_participant	Gitcoin info about some contributors	119
experiment_passport_stamp	Passport info about some contributors	574
experiment_vote	Vote record	14562
ethereum_transfer	Ethereum transfers related to GR15 contributors	8949267
polygon_transfer	Polygon transfers related to GR15 contributors	10915653
zksync_transfer	Zksync transfers related to GR15 contributors	2329056
contract_tag	Smart contract tags, e.g. opensea, binance	594949

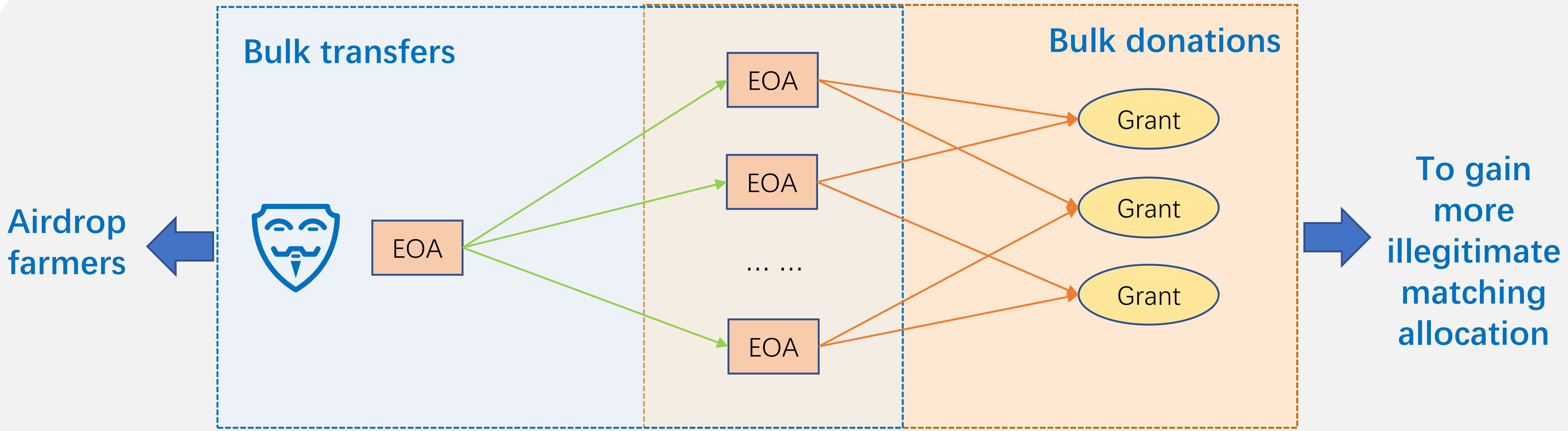
Statistics

– 2022-09-07 15:00:00 ~ 2022-09-22 23:59:59
– 4 Chains (Ethereum L1, zkSync L2, Polygon and Celo)
– 1440 Grants and 55585 Contributor Addresses
– 38 tokens (Eth, USDT, USDC, DAI, MATIC, WETH, etc)

Agenda

1. Introduction
    1. Team
    2. Deliverables (github & demo video)
2. Our Works
    1. Data Preparation
    2. Topic 1：Bulk Transfers & Donations
    3. Topic 2： Sequential Behavior Pattern Mining
    4. Topic 3 ： Asset-Transfer Graph Mining
    5. Topic 4: Grant Fraud
3. Summary, Suggestion and Future Works

Bulk Operations: Transfers & Donations

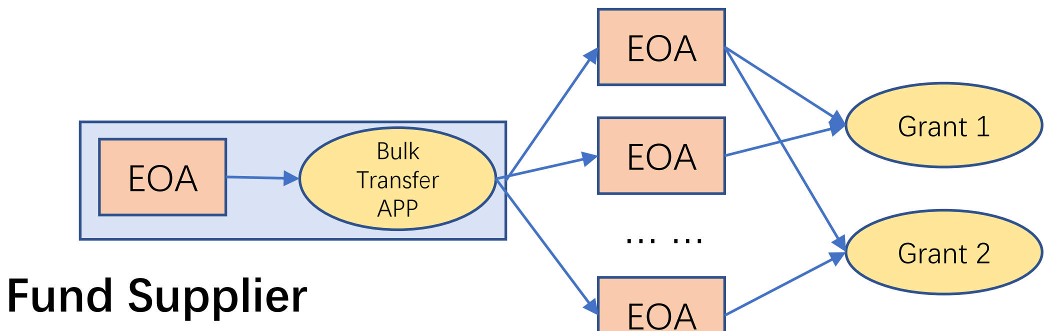


On Gitcoin platform, the sybil attack means that a user spreads their funds across multiple addresses and makes donation to the same project(s). We propose to detect Sybils by examining

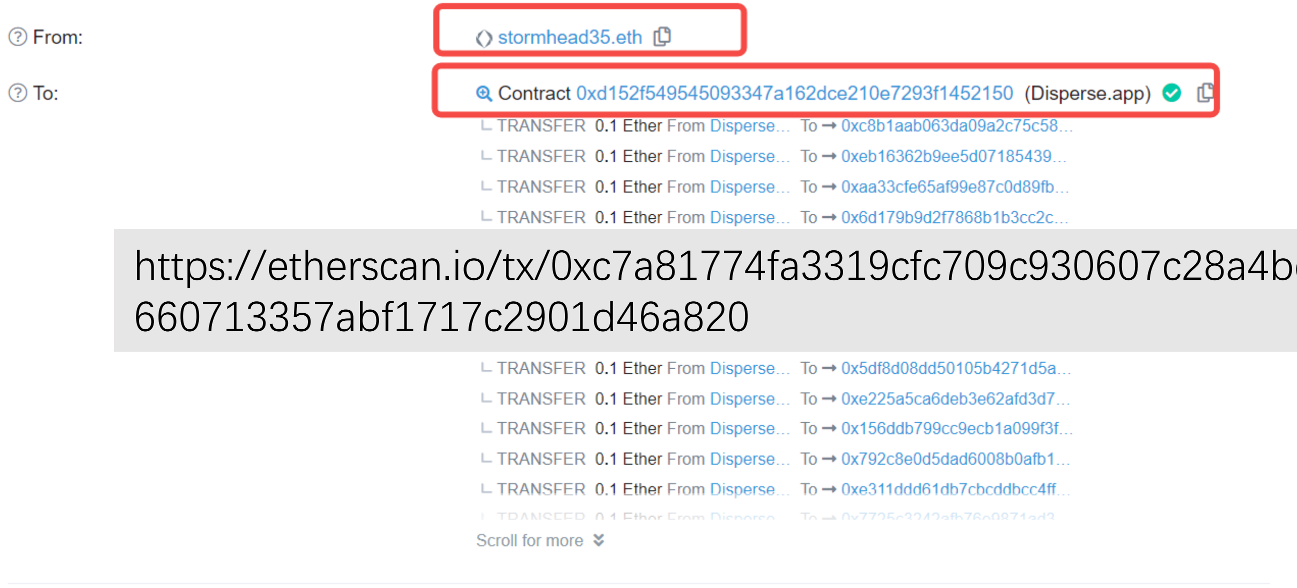
1. (Bulk Transfers) how the user spreads funds in a bulk way
2. (Bulk Donations) In what pattern they make donation to the same Grant(s)

Bulk Transfers Mining

Bulk transfer for fund preparations



An Real Example



https://etherscan.io/tx/0xc7a81774fa3319cfc709c930607c28a4bc56660713357abf1717c2901d46a820

Goal

1. Have a better understanding of What tools and How the sybils use for their fund preparations
2. Collect all the related addresses and sort them in terms of their sybils risks

Mining Pipeline

Select Bulk Transactions & Transfers → Design and Compute Risk Indicators (RIs) → Design and Compute Sybil Scorecard from RIs

Mining Pipeline

Select Bulk Transaction & Transfers → Design and Compute Risk Indicators (RIs) → Design and Compute Sybil Scorecard from RIs

Txn_hash	From	Time	Transfers_List
XXX	XXX	XXX	{(to_addr1,amount1,) (to_addr2,amount2,) (to_addr3,amount3,) …}

… …

A bulk transaction contains more than one transfers where all transfers have an identical sender address and every transfer may have different recipient address.

So, We select Bulk Transfers in a table where

(1) Every row corresponds to a bulk transaction
(2) The single sender is recorded, as well as the time
(3) A list of transfers where every transfer is represented by the to_addr, and amount

Mining Pipeline

Select Bulk Transaction & Transfers → **Design and Compute Risk Indicators (RIs)** → Design and Compute Sybil Scorecard from RIs

The Six Risk Indicators

- NumOfContr: The number of contributors in a transaction
- ContrRatio: The number of contributors divided by the total number of distinct recipients in the transaction
- ContrAmountRatio: The total amount of donations divide by the total amount of transfers-in

the greater the indicator, the greater the risk

- NumOfDistinctAmount: The number of distinct amount of transfers in a transaction
- MaxAmount: The maximum amount of the transfers in the transaction
- GapDay: The time difference between the transfer and donation

the greater the indicator, the less the risk

Mining Pipeline

Scorecard Results:

1. There are 21045 contributor addresses related to bulk transfers.
2. Based on the 6 RIs, we manually designed a scorecard, and the score distribution is

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13
NumOf Addr	9	689	1189	1841	3428	3703	3677	2436	1604	1592	537	256	71	13

3. We suggest that
   - Score > 9 (705 addresses) , Risk = High
   - Score>7 and < 9, Risk = Medium
   - Score < 9, Risk = Low;

Example 1

Addr	Txn_hash	NumOfContr	NumOfDistinctAmount	MaxAmount	ContrRatio	GapDay	ContrAmountRatio	Score
0xf4d84adf	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.339129489	11
0x2f3e3824	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.110870539	11
0x8dd578e	0xceab15a8l	37 out of 39	1	0.05	0.948717949	1	0.165388424	10
0xc8609af2	0xceab15a8l	37 out of 39	1	0.05	0.948717949	1	0.152663807	10
0xc7bf129l	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.179150249	11
0x0e0e131	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.083176408	11
0x0e0e131	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.083176408	11
0x21f72692	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.169564745	11
0x4a3fe89c	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.166366635	11
0x329b45e	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.169564745	11
0xd9ea00a	0xceab15a8l	37 out of 39	1	0.05	0.948717949	0	0.060138161	9
0x4f64a42l	0xceab15a8l	37 out of 39	1	0.05	0.948717949	1	0.082694212	10
0x4f64a42l	0xceab15a8l	37 out of 39	1	0.05	0.948717949	1	0.082694212	10
…	…	…	…	…	…	…	…	…

Reasoning:

- The transaction has 39 recipients, and 37 of them contributed to Gitcoin GR15. We report the 37 addresses out of 39 as a sybil cluster.
- In the same transaction, they equally received 0.05ETH from the same sender. The transaction are called by the disperse app.
- At the same day or one day later, they contributed to GR15.
- According to the 6 Ris, the Score is mostly 10 and 11 indicating very high likelihood to be sybil

ClusterID: 32
ClusterSize: 37
Discovered by: bulk transfer
RiskLevel： High

Some Noteworthy Bulk Transfer Apps (1)

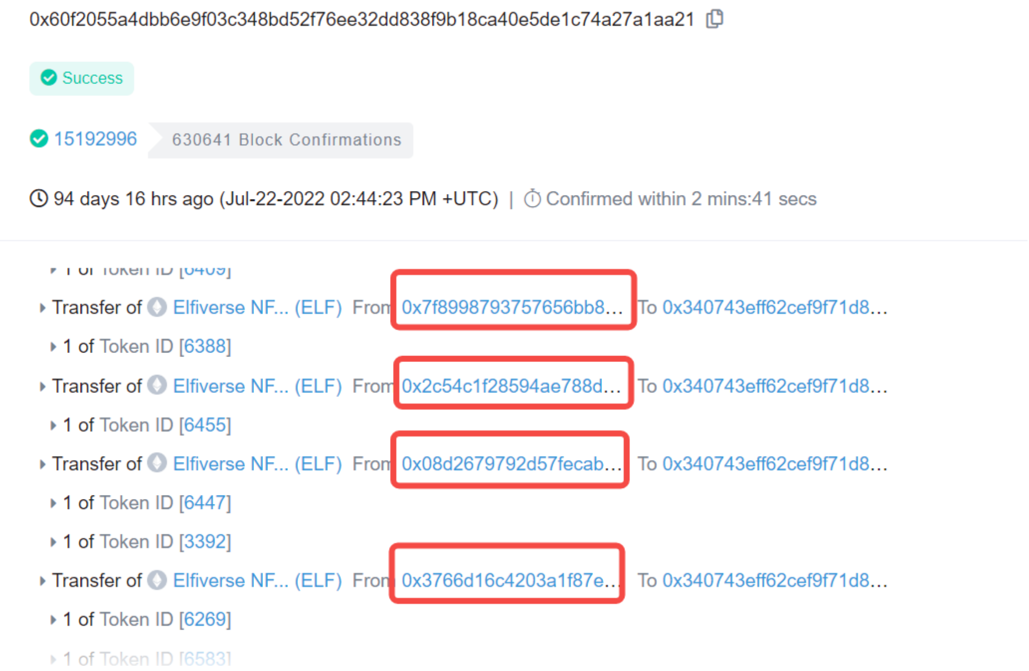
App	Smart Contract	Website	Description
multisender	'0xa5025faba6e70b84f74e9b1113e5f7f4e7f4859f'	https://multisender.app/	Send ERC20 Token or ETH to thousands of addresses out in 1 single transaction with Token Multisender.
disperse	'0xd152f549545093347a162dce210e7293f1452150'	https://disperse.app/	Distribute ether or tokens to multiple addresses
bulksender	'0xd1917932a7db6af687b523d5db5d7f5c2734763f'	https://bulksender.app/	Token bulksender
aztec_v2	'0xff1f2b4adb9df6fc8eafecdcbf96a2b351680455'	https://aztec.network/	The programmable privacy layer for web3
across_v2	0x4d9079bb4165aeb4084c526a32695dcfd2f77381	https://across.to/	Across is a cross-chain bridge that prides itself on its speed, security and low fees.
gnosis_safe	'0x0094477dfd27b9d5dc7ba610f26f0dd4ae64db5b', '0x81b2e8b475295f4254a38433b6739efe270fc88b', '0xb32aebf09cb331f853536b4370be8acf2d886775', '0xa788e30d0cd4d15f2159c686ff2ce8cf4be2c125', '0xa653ecfdd7987dd9b6bc284c3abd22bdb199159c'	https://gnosis.io/safe/	Smart contract-based multisig wallet

It is not necessary for an address to be a sybil if it uses these applications, but it is an importance signal.

Some Noteworthy Bulk Transfer Apps (2)

NFT marketplace Apps

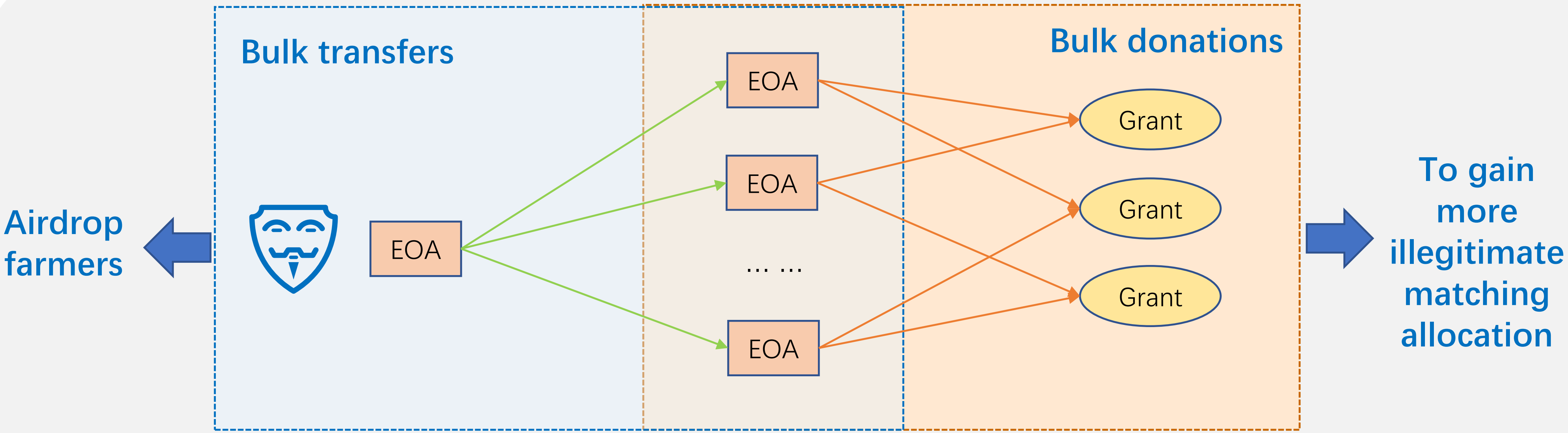
- opensea&seaport, element_ex, ragnarok (Looksrare)， async.art, rarible. boredapeyogaclub_v2
- Bulk NFT trades cause bulk fund transfers to contributors.



https://etherscan.io/tx/0x60f2055a4dbb6e9f03c348bd52f76ee32dd838f9b18ca40e5de1c74a27a1aa21

The contributors sold the same NFTs in one transaction. And then received funds from the same buyer simultaneously.

Bulk Operations: Transfers & Donations



On Gitcoin platform, the sybil attack means that a user spreads their funds across multiple addresses and makes donation to the same project(s). We propose to detect Sybils by examining

1. (Bulk Transfers) how the user spreads funds in a bulk way
2. (Bulk Donations) In what pattern they make donation to the same Grant(s)

Bulk Donations Mining: Data and Idea

Table containing detailed contribution Information

Grant	txn_id	txn_hash	Time	Token	Amount	......
Address 1						
Address 2	Table has the detailed contribution info with respect to a grant, such as the contributor address, txn ID and txn hash, the time of donation and the token and amount of the donation etc.					
Address 3						
......						
Address m						

The Idea

Attackers (or farmers) invest their time (to manipulate) and money (to donate) to perform Sybil attacks. Definitely they prefer to have a higher ROI (Return on investment).

Here are some assumptions as to greedy Sybils:

1. They donate the same grant
2. They donate as small amount as possible
3. They make use of script/tools to donate with the same parameter setting such as ChainID, Layer of Chain, token, amount
4. They donate in a sequential way, very closely

Indicators(Or Variables) and Scores

We encode our plain English assumptions as filtering variables

Var3：Group1Proportion
NumOfAddresses in Group1 divided by the total number of addresses in the grant

Var6：Q3TimeDifferences
The 3rd quartile (Q3) of the time differences between successive donations

Var7：CVTimeDifferences
The coefficient of variation of the time differences between successive donations

Var1：Amount

Var2：#ContributionsGroup1
NumOfContributions in Group1

Var5：Group2Proportion
NumOfAddresses in Group2 divided by the total number of addresses in the grant

Var4：#ContributionsGroup2
NumOfContributions in Group2

Grouping Criteria 1：

- Same Grant
- Same Network
- Same Amount
- Same Chain
- Same Token

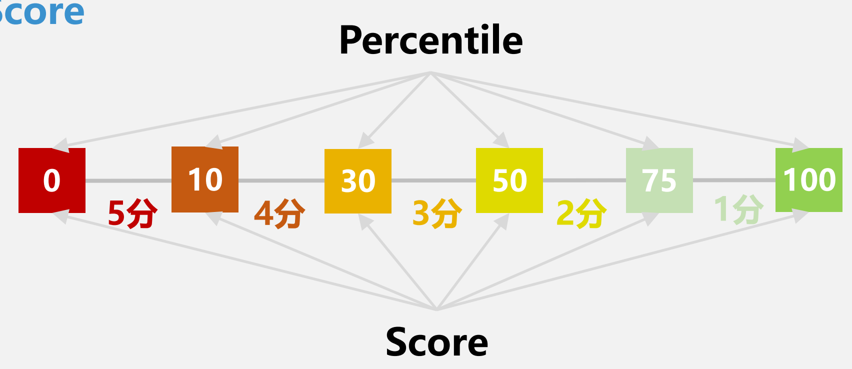
Grouping Criteria 2：

Same grouping criteria as of grouping 1

- Time differences between two successive donations are <=30mins

Var as Risk Score

Var2
Var3
Var4
Var5
Var6
Var7



Var1 (Amount)： < 1U：5分；>1U and <=1.1U：4分；=1U：3分；>1.1U and <=1.3U：2分；>1.3U：1分

Scoring donation and address

Donation Risk Score:

$$S_{donation} = \sum_{i=1}^{7} S_i$$

Address Risk Score:

$$S_{address} = \text{MAX}\ (S_{donation\ 1}, \dots \dots, S_{donation\ n})$$

Where **donation 1** to **donation n** are donated by current address.

Results and Risk Level

Score Distribution

Trans Score	Trans CNT	Address CNT	Trans PCT	Address PCT
30	351	132	0.1%	0.2%
29	240	94	0.1%	0.2%
28	1,343	520	0.3%	0.9%
27	2,461	818	0.6%	1.5%
26	3,002	869	0.8%	1.6%
25	4,722	1,484	1.2%	2.7%
24	5,904	1,769	1.5%	3.2%
23	6,665	1,880	1.7%	3.4%
22	7,367	1,825	1.8%	3.3%
21	7,523	1,656	1.9%	3.0%
20	8,479	1,867	2.1%	3.4%
19	9,383	2,190	2.4%	4.0%
18	8,003	1,783	2.0%	3.2%
17	7,359	1,587	1.8%	2.9%
16	7,016	1,589	1.8%	2.9%
15	6,925	1,352	1.7%	2.4%
14	5,778	1,114	1.4%	2.0%
13	4,567	739	1.1%	1.3%
12	3,177	536	0.8%	1.0%
11	1,872	386	0.5%	0.7%
10	1,159	225	0.3%	0.4%
9	463	89	0.1%	0.2%
8	46	18	0.0%	0.0%
7	36	3	0.0%	0.0%
0	295,373	30,778	74.0%	55.7%

*** Score=0 are donations not likely to be sybils*

Risk Levels

- **High: Score >= 21** — #OfAddress: **14005, 25.3%**
- **Medium: Score < 21 and Score > 0** — #OfAddress: **16040, 29.0%**
- **Low: Score = 0** — #OfAddress: 25258, 45.7%

** Note: the total number reported also includes results of another thread on bulk donations

Example 1: Sybils Manipulated by the Same UserID

The 10 Sybil Donations

address	time	time_gap	chain	token	amount
0x7edb5ed01fd0c42ea3273fc1cb1f8943b12978ad	2022-09-21 11:26:39	0.42	eth_zksync	DAI	1.4
0xa94110480a2 [illegible]	2022-09-21 11:27:04	0.37	eth_zksync	DAI	1.4
0xca5d94ab99 [illegible]	2022-09-21 11:27:26	0.40	eth_zksync	DAI	1.4
0x46dbaecb16 [illegible]	2022-09-21 11:27:50	0.37	eth_zksync	DAI	1.4
0x350d84d0ed [illegible]	2022-09-21 11:28:12	0.37	eth_zksync	DAI	1.4
0xcd2519d2f3 [illegible]	2022-09-21 11:28:34	0.40	eth_zksync	DAI	1.4
0xfe565cd155 [illegible]	2022-09-21 11:28:58	0.38	eth_zksync	DAI	1.4
0x6f3d616590 [illegible]	2022-09-21 11:29:21	0.42	eth_zksync	DAI	1.4
0x101bec121c54 [illegible]	2022-09-21 11:29:46	0.62	eth_zksync	DAI	1.4
0xd704bf5cc05ff8465903da4c515eb223b5f2eb4f	2022-09-21 11:30:23		eth_zksync	DAI	1.4

Grant ID: 7202
Cluster ID: 77809
Cluster Size: 10
Discovered by: bulkdonation
RiskLevel: High

Some Normal Donations

time	time_gap	chain	token	amount
2022-09-08 05:52:59	41.87	eth_std	DAI	1.16
2022-09-08 06:34:51	38.43	eth_zksync	DAI	25.00
2022-09-08 07:13:17	100.77	eth_zksync	USDC	1.22
2022-09-08 08:54:03	57.52	eth_std	ETH	1.15
2022-09-08 09:51:34	47.07	eth_std	DAI	1.05
2022-09-08 10:38:38	94.30	eth_zksync	ETH	1.31
2022-09-08 12:12:56	115.27	eth_polygon	USDC	1.10
2022-09-08 14:08:12	40.33	eth_zksync	ETH	1.14
2022-09-08 14:48:32	20.70	eth_polygon	MATIC	1.25
2022-09-08 15:09:14	29.63	eth_zksync	DAI	1.02
2022-09-08 15:38:52		eth_std	ETH	32.64

Reasoning

The same Gitcoin User (UserID= e18388e14838506df27f901c8b62de6c4fd6c Da5b5e332e955b078fe6482bc96) made 10 donations to GrantID=7202 with 10 wallets. These happened in 4 mins, with a very close time interval = 30 secs between every two successive donations. Every wallet contributed 1.4U DAI on zkSync.

On the contrary, normal donations do not have apparent pattern in token aggregation, amount aggregation and chain aggregations. The time intervals between every successive donations are not all small.

Example 2: Donations with Same & Small Amounts

Reasoning

Within about 2 hours, there are 61 donations to Grant 6713. The time intervals are as small as 30 secs to 1 minute. Every addresses donate 0.086U MATIC on polygon.

address	time	time_gap	chain	token	amount
0x119b9dbe4a5e3fae94fce23511562a0dd78cb6d3	2022-09-15 12:18:32	2.25	eth_polygon	MATIC	0.086
0xf91cd9bfc08d1df946138fa582428491d2a78778	2022-09-15 12:20:47	1.50	eth_polygon	MATIC	0.086
0x930df074acf694238bdf6e1c947e1c5442d5f019	2022-09-15 12:22:17	1.27	eth_polygon	MATIC	0.086
0x2da181eda79285e810ebcab3eec32c706aa28a9c	2022-09-15 12:23:33	1.23	eth_polygon	MATIC	0.086
0x5df8d08dd50105b4271d5ad07d094ace85cf0ee6	2022-09-15 12:24:47	2.00	eth_polygon	MATIC	0.086
0x6d179b9d2f7868b1b3cc2cd5b3e93eb0eb9d1174	2022-09-15 12:26:47	0.98	eth_polygon	MATIC	0.086
0x792c8e0d5dad6008b0afb1c5a4565b160a965485	2022-09-15 12:27:46	3.35	eth_polygon	MATIC	0.086
0xe311ddd61db7cbcddbcc4ffb8ac68eff6b7820df	2022-09-15 12:31:07	2.45	eth_polygon	MATIC	0.086
0x7725c3242afb76e9871ad3e7c879f4c6d5db0092	2022-09-15 12:33:34	1.30	eth_polygon	MATIC	0.086
0x5ff1dd[illegible]665b7e	2022-09-15 12:34:52	1.13	eth_polygon	MATIC	0.086
0x5ba36[illegible]e03d3	2022-09-15 12:36:00	1.17	eth_polygon	MATIC	0.086
0x1e6c3c[illegible]836501	2022-09-15 14:33:08	0.25	eth_polygon	MATIC	0.086
0x76025c1ebb7c6d61190761d2985c77cb2f464aae	2022-09-15 14:33:23	0.27	eth_polygon	MATIC	0.086
0xb251694f56777115ca3bc840e36f9cb1581f0883	2022-09-15 14:33:39	0.27	eth_polygon	MATIC	0.086
0x844203780009ed82a1d2c9c220f71f88091f018e	2022-09-15 14:33:55	0.35	eth_polygon	MATIC	0.086
0xa9174b240ea4001c7b228916ce34791b111e6fbb	2022-09-15 14:34:16	3.50	eth_polygon	MATIC	0.086
0x43a234c2c9f698d7dc56cfb32b7dfac31f7fe0de	2022-09-15 14:37:46		eth_polygon	MATIC	0.086

Grant ID: 6701
Cluster ID: 250829
Cluster Size: 61
Discovered by: bulkdonation
RiskLevel： High

Agenda

1. Introduction
   1. Team
   2. Deliverables (github & demo video)
2. Our Works
   1. Data Preparation
   2. Topic 1：Bulk Transfers & Donations
   3. Topic 2： Sequential Behavior Pattern Mining
   4. Topic 3 ： Asset-Transfer Graph Mining
   5. Topic 4: Grant Fraud
3. Summary, Suggestion and Future Works

Sequential Behavior Pattern Mining

Possible On-Chain Activities					
Transfer Asset	Interact with SC	Mint NFT	Claim Airdrop	Donate Gitcoin	…

In the form of transactions, event, log, internal call

Account transactions https://zkscan.io/explorer/accounts/0x17bcb22e0ce414b633af35be8d62f26f588247c8

Tx Hash	Type	Amount	From	To	Created
0xbc4c066565...	Transfer	0 ETH	0x17bcb22e0c...	0x17bcb22e0c...	about 1 month ago
0x3517c55873...	Transfer	0.0007 ETH	0x17bcb22e0c...	0x3ccb0f1f6c...	about 1 month ago
0x0ff2413bcc...	Transfer	0.0007 ETH	0x17bcb22e0c...	0x99b36fdbc5...	about 1 month ago
0x051fc44f5a...	Transfer	0.0007 ETH	0x17bcb22e0c...	0x18aa467e40...	about 1 month ago
0x8bc88b17b4...	Transfer	0.0014 ETH	0x17bcb22e0c...	0x000002c34b...	about 1 month ago
0xbcfbc8ff43...	ChangePubKey		0x17bcb22e0c...		about 1 month ago
0xcbbecf1be7...	Deposit	0.11 ETH	L1 0x17bcb22e0c...	0x17bcb22e0c...	about 1 month ago

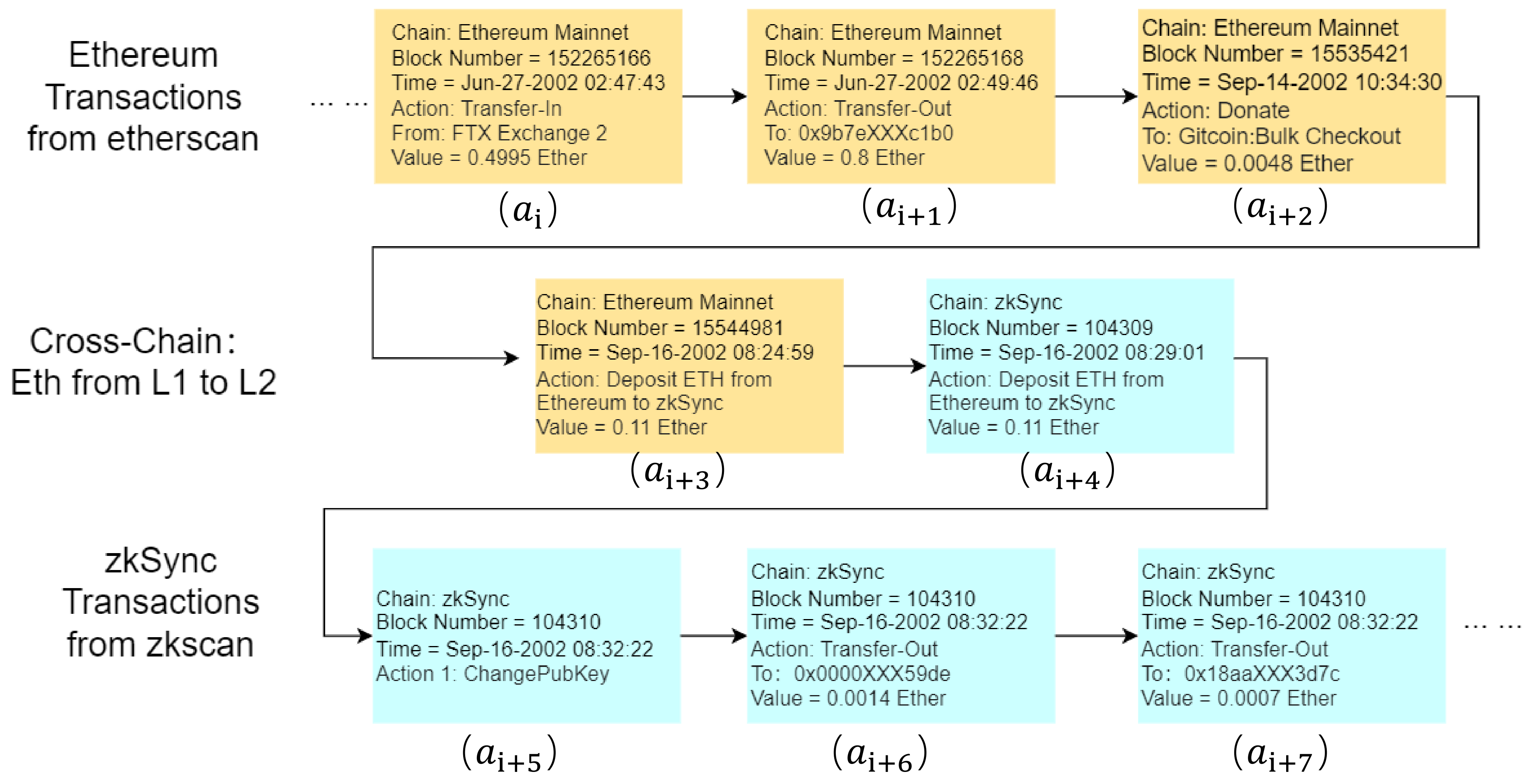
Latest 12 from a total of 12 transactions https://etherscan.io/address/0x17bcb22e0ce414b633af35be8d62f26f588247c8

Txn Hash	Method	Block	Age	From		To	Value	Txn Fee
0xcbbecf1be7f31563d3b...	Deposit ETH	15544981	33 days 19 hrs ago	0x17bcb22e0ce414b633...	OUT	zkSync	0.11 Ether	0.00045689
0xdd0854be06da9e9749...	Donate	15535421	35 days 4 hrs ago	0x17bcb22e0ce414b633...	OUT	Gitcoin: Bulk Checkout	0.0048 Ether	0.00059183
0xe375690bc34992972a...	Transfer	15225168	84 days 12 hrs ago	0x17bcb22e0ce414b633...	OUT	0x9b7e57b8f675a954d9...	0.8 Ether	0.00035345
0xaed7daa09cce0dec22...	Transfer	15225161	84 days 12 hrs ago	FTX Exchange 2	IN	0x17bcb22e0ce414b633...	0.4995 Ether	0.00045926

Make EOA(address) behavior computable so as to facilitate sybil discovery

- Sequential Behavior Representation
- Sequential Behavior Similarity Computation
- Usecase 1 Searching
- Usecase 2 Clustering

Sequential (Cross-Chain) Behavior Representation



Similarity Definition

Mathematically, for EOA1's behavior $s_1 = (a_1, a_2, ...a_n)$ and EOA2's behavior $s_2 = (b_1, b_2, ...b_m)$, define that

> $sim(s_1, s_2) = 1$ if and only if
>
> (1) $n = m$ and
>
> (3) for every pair of actions a_i, b_i , they are identical with only negligible time difference.

More sophisticated metrics can be experimented with in the future.

Use Case 1: Searching

Hi Gitcoin. Watch out!

Here is a suspicious contributor

0x0aa599d8a9a1
418dc2df0bf240
95f7d87c0b1929

{Chain:ETH； Blocknumber:15392994；
time:2022-08-23 06:47:00；
to: 0xd93206bd0062cc054e397ecccdb8436c3fa5700e；
Action: 0x42842e； value: 0.0}

{ Chain:ETH； Blocknumber:15392994；
time： 2022-08-23 06:47:00；
to： 0xd93206bd0062cc054e397ecccdb8436c3fa5700e；
Action： 0x42842e； value： 0.0}

{ Chain:ETH； Blocknumber:15392994；
time： 2022-08-23 06:47:00；
to： 0xd93206bd0062cc054e397ecccdb8436c3fa5700e；
Action： 0x42842e； value： 0.0}

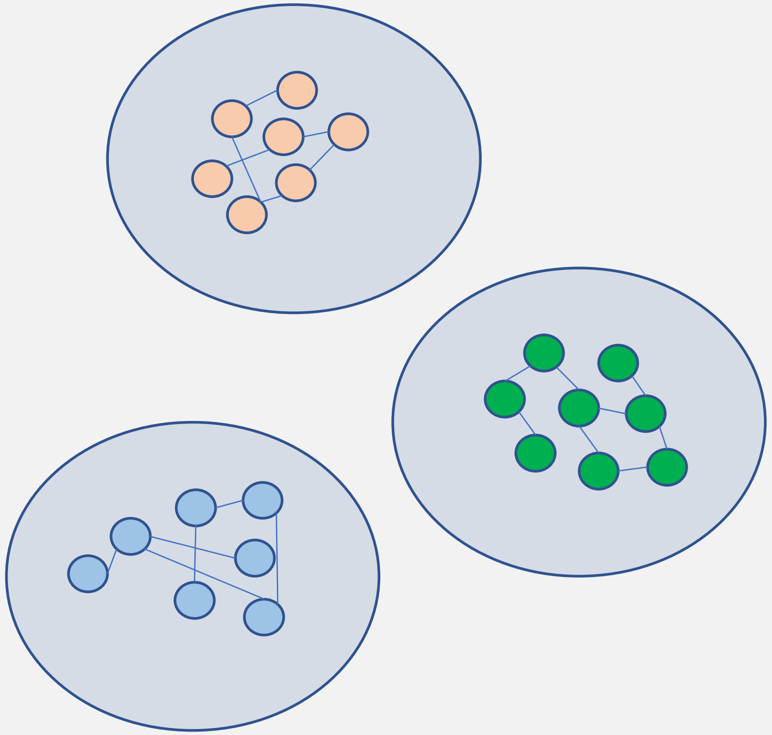
{ Chain:ETH； Blocknumber:15392994；
time: 2022-08-23 06:47:00；
to： 0xd93206bd0062cc054e397ecccdb8436c3fa5700e；
Action:0x42842e； value： 0.0}

{Chain:ETH； Blocknumber:15392994；
time： 2022-08-23 06:47:00；
to： 0xd93206bd0062cc054e397ecccdb8436c3fa5700e；
Action: 0x42842e； value： 0.0}

{Chain:ETH； Blocknumber:15392994；
time： 2022-08-23 06:47:00；
to： 0xd93206bd0062cc054e397ecccdb8436c3fa5700e
Action： 0x42842e； value： 0.0}

INPUT an address, OUTPUT the list of addresses having almost the same on-chain behaviors

Use Case 2: Clustering



1. Sequential Behavior Representation
2. Similarity Definition
3. Clustering： A variety of distance-based algorithms such as spectral clustering, hierarchical clustering can be used. Here, Connected Component Based Clustering,
    1. If (A,B) in a cluster; (B,C) in a cluster, then
    2. (A,B,C) are all in the same cluster

    Result in 115 clusters with size >=5

Cluster Size	[2,5)	[5,10)	[10,20)	[20,30)	[30,40)	[40,50)	[50,60)	=75	=192	>=5 Total
# of CLusters	380	74	19	8	6	3	3	1	1	115 Clusters (1669 Sybils)

Example 1

The Cluster of Size = 75	# of Actions
1	4
2	4
3	4
…	
…	
…	
…	4
74	4
75	4

ClusterID: 32
ClusterSize: 75
Discovered by: behavior
RiskLevel: High

Reasoning

All these addresses has only 4 actions. Apparently from etherscan, the four actions are

127 days ago
Transfer in from Binance

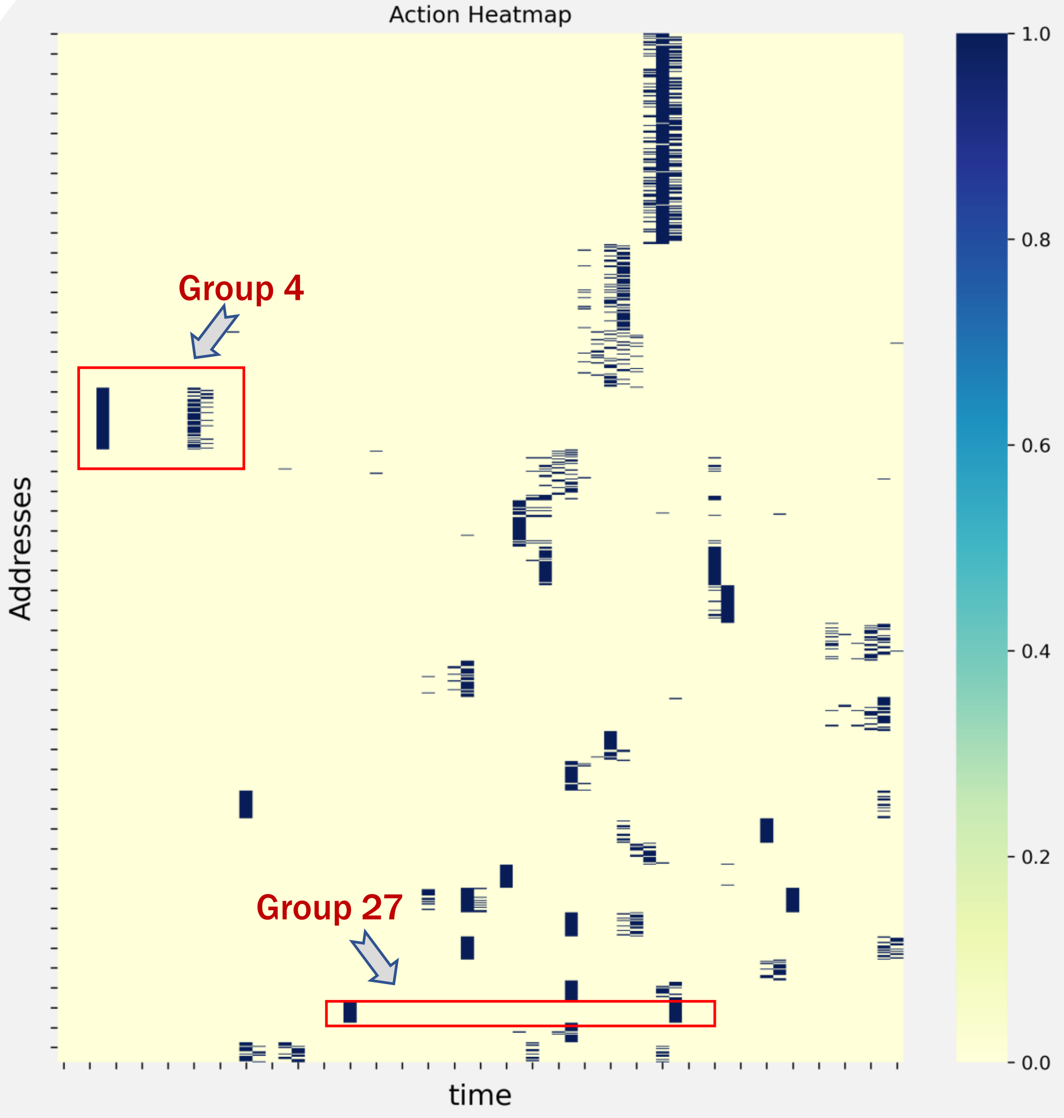
127 days ago
Deposit ETH to Arbitrum

127 days ago
Send to L2 via HOP Ethereum Bridge

41 days ago
Donate to Gitcoin: Bulk Checkout

Visualization and Examinations

Heatmap visualization to represent clusters behaving in a similar way



Group 4: 56 contributors

1. 20220907, received 0.06Eth from 56 addresses
2. 20220909 11am-15pm, Deposit L2 with 0.055Eth
3. On zkSync, they all
    1. Mint NFT
    2. Swap some ETH for DAI
    3. Donate GR15

Group 27: 19 contributors

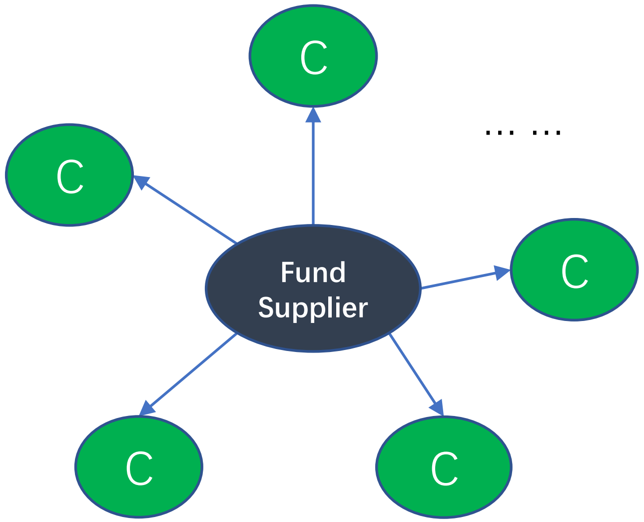
1. 20220912, received 0.06Eth for the 1st time
2. 20220912 Deposit L2 with 0.055Eth
3. On zkSync, they all
    1. Mint NFT
    2. Swap the same amount ETH for DAI
    3. Donate GR15
4. 20220924 On L1, Interact with Socket:Registry

Agenda

1. Introduction
    1. Team
    2. Deliverables (github & demo video)
2. Our Works
    1. Data Preparation
    2. Topic 1： Bulk Transfers & Donations
    3. Topic 2： Sequential Behavior Pattern Mining
    4. Topic 3 ： Asset-Transfer Graph Mining
    5. Topic 4: Grant Fraud
3. Summary, Suggestion and Future Works

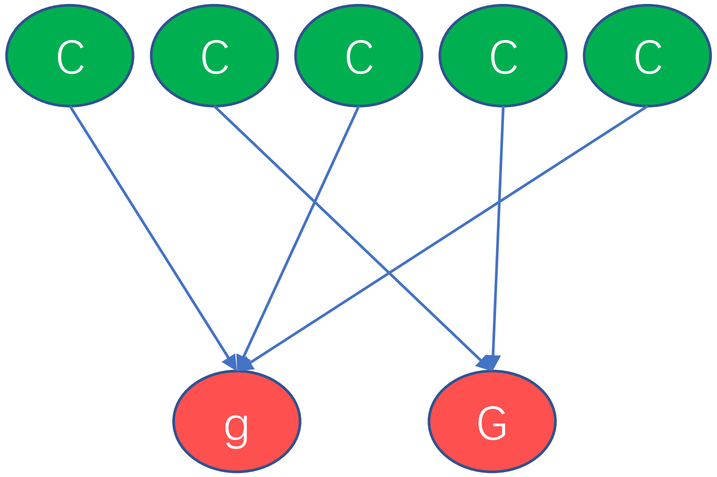
Asset-Transfer Graph (ATG)

Asset-Transfer Graph (ATG) represents the relationship between Contributors, Grants(Gitcoin), Fund Suppliers and Residual Collectors in terms of their asset (ETH, MATIC and ERC20 tokens) transfers.



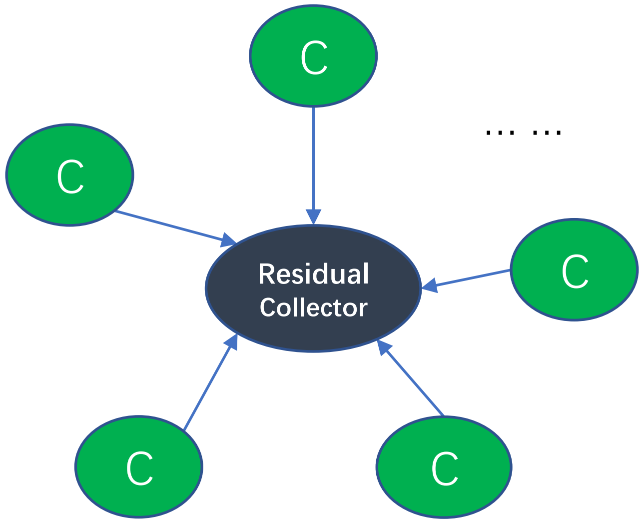
Phase1: Fund Preparation

Fund Supplier transfers funds to 1 or more contributors



Phase 2: Donation

Contributors transfer fund to grant Gitcoin to make donations



Phase 3: Residual Collection

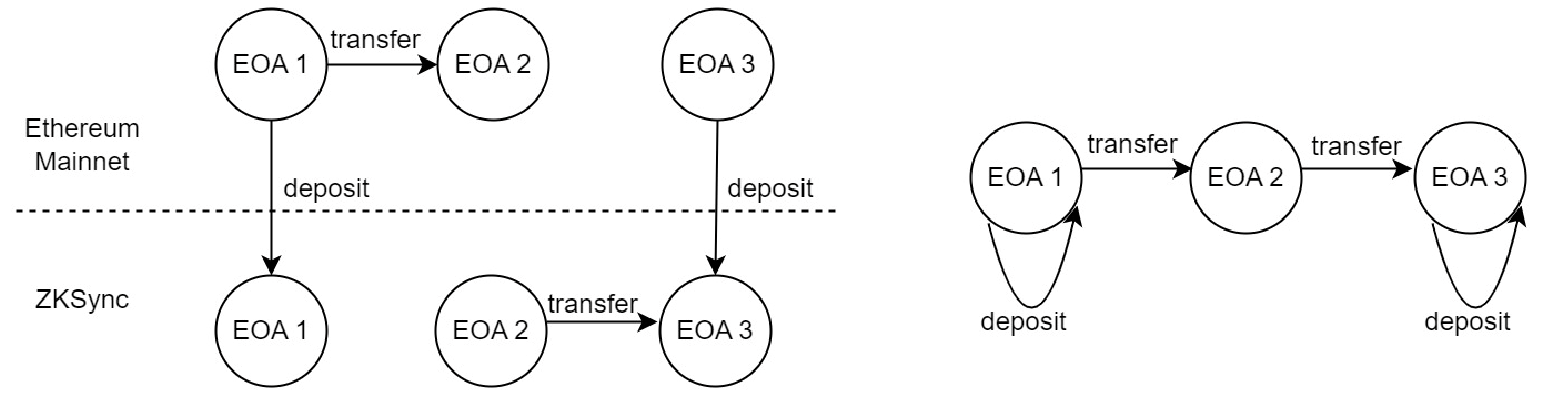
After donation, contributors transfers residual fund to residual collector

Overall Graph Constructions

Asset-Transfer: Transfers from A → B where A OR B is a contributor

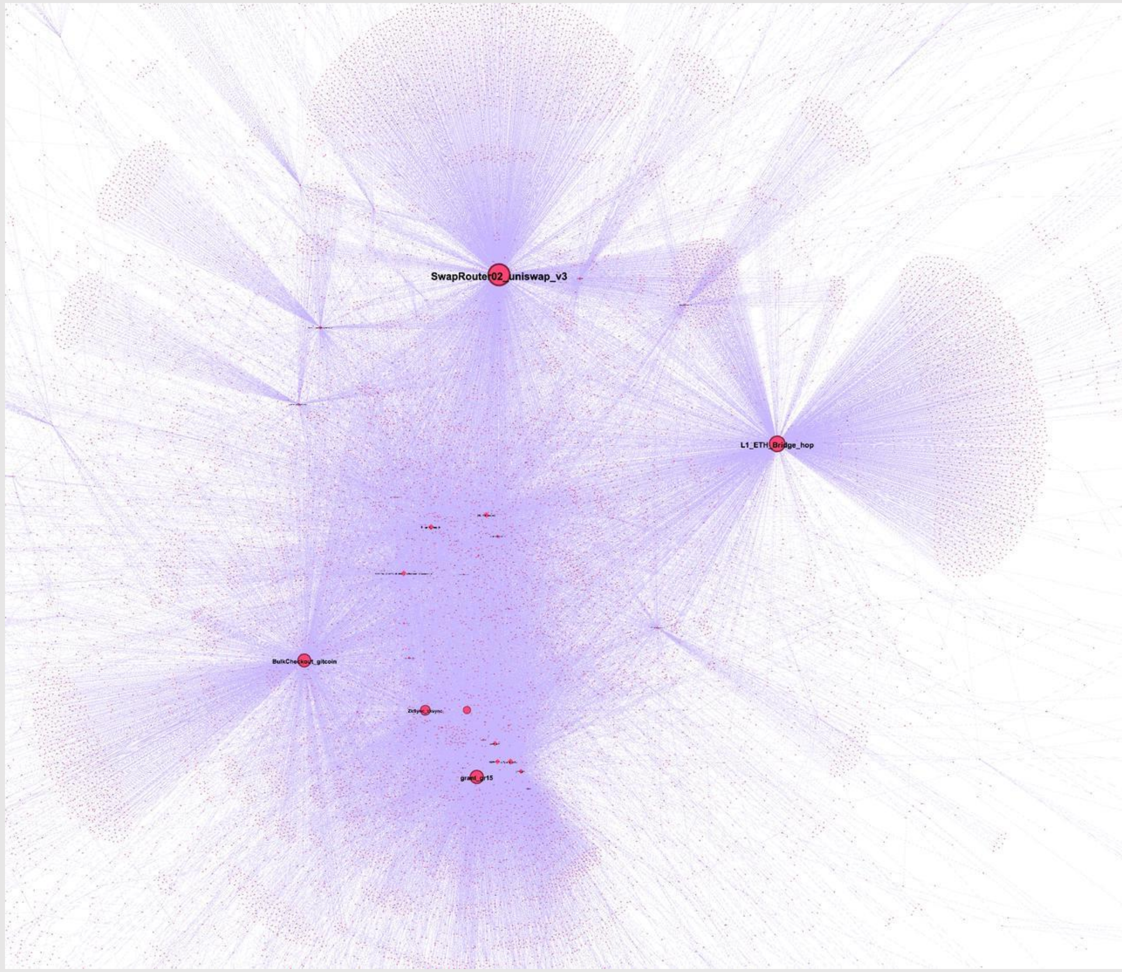
1. Include the Assets = ETH, USDC, USDT, DAI on Ethereum, Polygon and zkSync
2. Date: Sept 1st - Oct 19th

ATGraph Size: 1.4M edges



The two graphs represnet the same asset transfers.
As shown, Ethereum mainnet and ZKSync transfers are aggregated and illustrated in one graph.

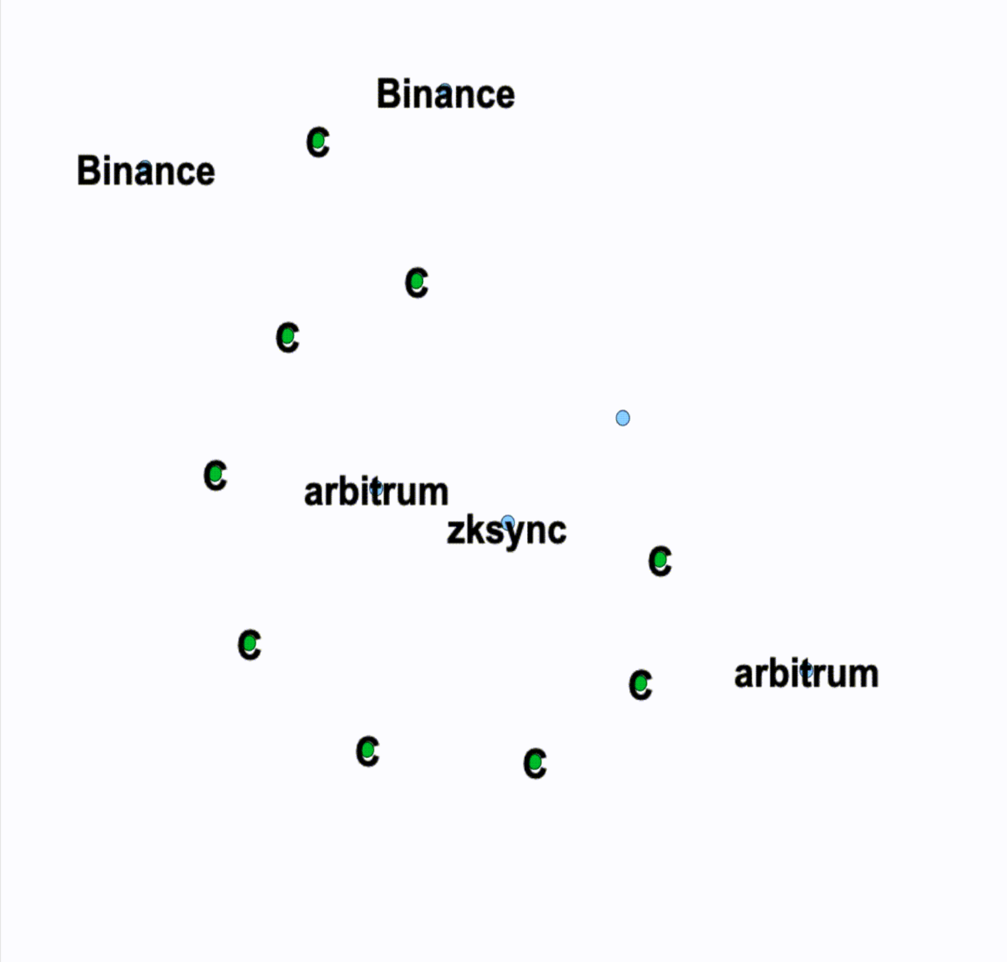
Overall Graph Visualization



Visualization using Gephi: The open graph viz platform

HUB	Label	Degree
0x7d655c57f71464b6f83811c55d84009cd9f5221c	BulkCheckout_gitcoin	52201
0x00000000006c3852cbef3e08e8df289169ede581	Seaport_seaport	33926
0xc098b2a3aa256d2140208c3de6543aaef5cd3a94	FTX Exchange 2_	30229
0xabea9132b05a70803a4e85094fd0e1800777fbef	ZkSync	17618
0x283af0b28c62c092c9727f1ee09c02ca627eb7f5	ETHRegistrarController_3nameservice	17150
0x68b3465833fb72a70ecdf485e0e4c7bd8665fc45	SwapRouter02_uniswap_v3	17035
0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2	WETH9_zeroex	9765
0x4d9079bb4165aeb4084c526a32695dcfd2f77381	Ethereum_SpokePool_across_v2	9326
0x5bf5bcc5362f88721167c1068b58c60cad075aac	ProofOfStake_Pages_vitaliks_book	8919
0xff1f2b4adb9df6fc8eafecdcbf96a2b351680455	RollupProcessor_aztec_v2	8301
0x4dbd4fc535ac27206064b68ffcf827b0a60bab3f	Inbox_arbitrum	8198
0xb8901acb165ed027e32754e0ffe830802919727f	L1_ETH_Bridge_hop	6181
0x20f780a973856b93f63670377900c1d2a50a77c4	ERC721OrdersFeature_element_ex	5288
0xdef1c0ded9bec7f1a1670819833240f027b25eff	ExchangeProxy_zeroex	4684

Example 1: Chain-Like Attack



Account Balances

Token	Amount
DAI	0.4916672497
ETH	0.000667087933

Account transactions

Tx Hash	Type	Amount	From	To	Created
0xf7740371f8...	Transfer	0 DAI	0x7842dd0d3e...	0x7842dd0d3e...	about 1 month ago
0xb7e9074cff...	Transfer	0.055 DAI	0x7842dd0d3e...	0xde21f72913...	about 1 month ago
0x161e2f74a8...	Transfer	1.1 DAI	0x7842dd0d3e...	0x10e87b05fe...	about 1 month ago
0x4ee2203d0f...	Swap	0.0007294769598 ETH → 0.9012441404 DAI	0x7842dd0d3e...	0x5f8e570ce0...	about 1 month ago
0x7c2dc5884a...	Transfer	0 DAI	0x7842dd0d3e...	0x7842dd0d3e...	about 1 month ago
0x1bbdc65dd7...	Transfer	0.055 DAI	0x7842dd0d3e...	0xde21f72913...	about 1 month ago
0x788e9b9a9e...	Transfer	1.1 DAI	0x7842dd0d3e...	0x99b36fdbc5...	about 1 month ago
0x6e3d247aa0...	Swap	0.0014266351072 ETH → 2.0043231093 DAI	0x7842dd0d3e...	0x9c3b23d57a...	about 1 month ago
0x6153299c3c...	ChangePubKey		0x7842dd0d3e...		about 1 month ago
0x83bab6fb9f...	Deposit	0.003 ETH	L1 0x7842dd0d3e...	0x7842dd0d3e...	about 2 months ago

denote

swap for DAI

get ETH from L1 net

L1:

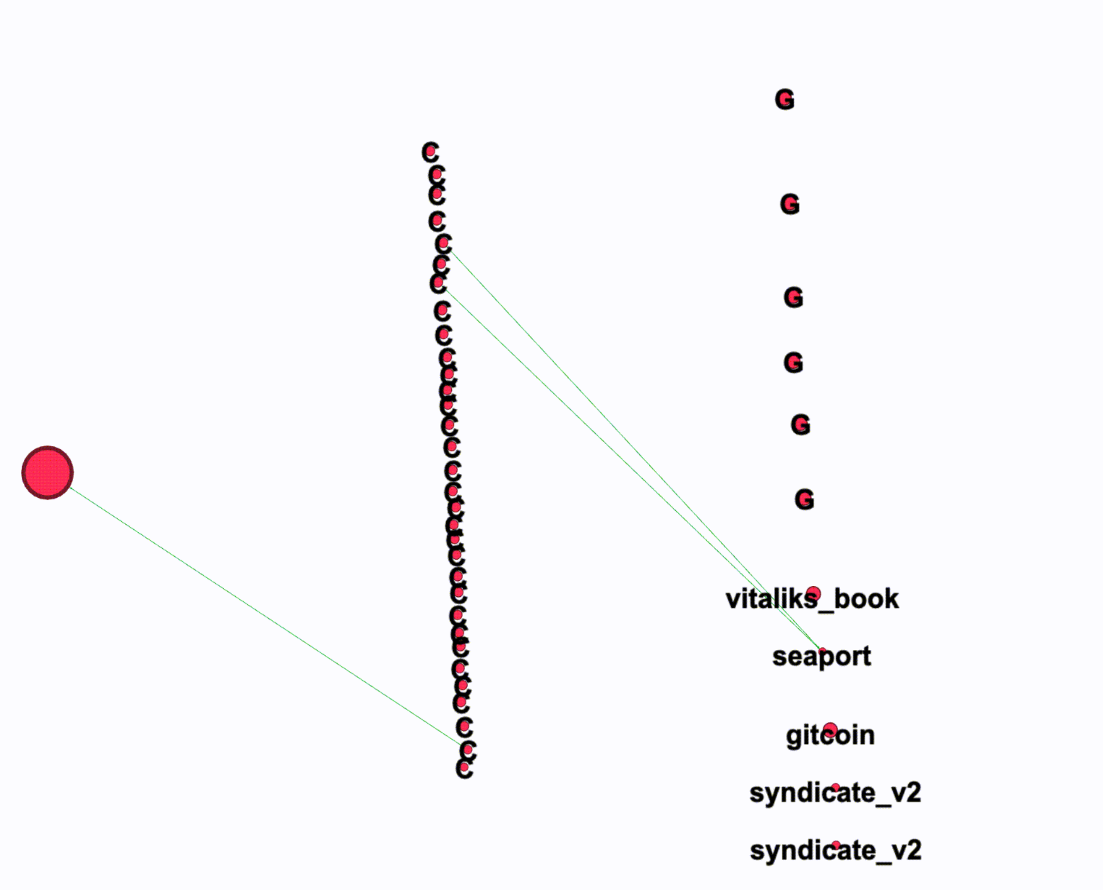
1. ETH from Binance to the first contributor
2. ETH goes through the 9 Contributors in turn

L2:

1. Deposit to L2, and then Swap to DAI
2. Donate

Example 2: Diamond Attack

15470285	2022-09-04 7:12:49	0x08b98ffc06b29fc17f65...	call	0x8a5149b12a333c1d6d...	→	0xb2e6ed0d632174d24a...	0.0003 Ether
15470279	2022-09-04 7:11:55	0xe93967caf0d479d0ae...	call	0x8a5149b12a333c1d6d...	→	0x89756931e8099bc5be...	0.0005 Ether
15470275	2022-09-04 7:10:29	0x3d562068e7a5c57de4...	call	0x8a5149b12a333c1d6d...	→	0xf008ea87d172fe6c53e...	0.00048 Ether
15470271	2022-09-04 7:09:11	0xb545e47186b47ed473...	call	0x8a5149b12a333c1d6d...	→	0xdb91d50690091a52ac...	0.0004 Ether
15470269	2022-09-04 7:08:35	0xaaa6742c3dceb90f10...	call	0x8a5149b12a333c1d6d...	→	22to22.eth	0.003 Ether
15470263	2022-09-04 7:07:28	0x68d40e702c5437d6f5...	call	0x8a5149b12a333c1d6d...	→	0x596e3b53a71e22f806...	0.00048 Ether
15470256	2022-09-04 7:05:56	0x716ce5e6a409b52ad4...	call	0x8a5149b12a333c1d6d...	→	0xa22d18176d63f73cb7...	0.0003 Ether
15470254	2022-09-04 7:05:16	0x5a46a14aed03ec7cbf...	call	0x8a5149b12a333c1d6d...	→	0xcfa0bdbb72563ca512...	0.0002 Ether
15470249	2022-09-04 7:04:05	0x88bb0ed5649251a6b6...	call	0x8a5149b12a333c1d6d...	→	0x9aeb1de021b7530c75...	0.0003 Ether
15470246	2022-09-04 7:03:30	0xb1171238c98de04212...	call	0x8a5149b12a333c1d6d...	→	0x924842dde58c84d153...	0.0004 Ether
15470233	2022-09-04 7:01:13	0xae7b7dab5d1ceee654...	call	0x8a5149b12a333c1d6d...	→	🌈nft🌈.eth	0.001 Ether
15470229	2022-09-04 7:00:54	0x86bb003b4282779c1f...	call	0x8a5149b12a333c1d6d...	→	0xdb6edad6efa9498973...	0.0003 Ether
15470225	2022-09-04 7:00:15	0xbae36ec6cb41121eb8...	call	0x8a5149b12a333c1d6d...	→	0x07078da8072f969017...	0.0005 Ether
15470221	2022-09-04 6:58:29	0x9949713480c6db1046...	call	0x8a5149b12a333c1d6d...	→	0xe3a113fd0c3d4e83ef6...	0.0004 Ether
15470216	2022-09-04 6:57:29	0x3c2a96221fd1d259d6...	call	0x8a5149b12a333c1d6d...	→	0x450a673ad966dfb327...	0.00044 Ether
15470213	2022-09-04 6:56:21	0xb90a48605687cbcdfe...	call	0x8a5149b12a333c1d6d...	→	0x11d9a85ac9a715f7d1...	0.0004 Ether
15470198	2022-09-04 6:53:53	0x810c74f68ea59f4cdd9...	call	0x8a5149b12a333c1d6d...	→	0xccc3ffbd9352dc82760...	0.00049 Ether
15470191	2022-09-04 6:52:35	0x4001e9088378b14c8a...	call	0x8a5149b12a333c1d6d...	→	0x43fa7ed9169a92ccf87...	0.00046 Ether
15470187	2022-09-04 6:50:41	0x34684b55b1e3aa98c5...	call	0x8a5149b12a333c1d6d...	→	0xd36d44060f58da177fa...	0.00048 Ether
15470182	2022-09-04 6:49:35	0xa1fa8d6bbd05854989...	call	0x8a5149b12a333c1d6d...	→	0x5bc9bb3bc89558a02b...	0.00044 Ether
15470169	2022-09-04 6:46:50	0x049ac2bbb25bf0d520...	call	0x8a5149b12a333c1d6d...	→	0x0ec6a2ae39a2df7dac...	0.00045 Ether
15470161	2022-09-04 6:45:22	0x9c378a01cad138590c...	call	0x8a5149b12a333c1d6d...	→	0xf4f1a96a0f800ad8dd3...	0.00051 Ether
15470156	2022-09-04 6:44:28	0x93e2787f5b25b7586e...	call	0x8a5149b12a333c1d6d...	→	0xb2a7cd2a4e97b18239...	0.00046 Ether



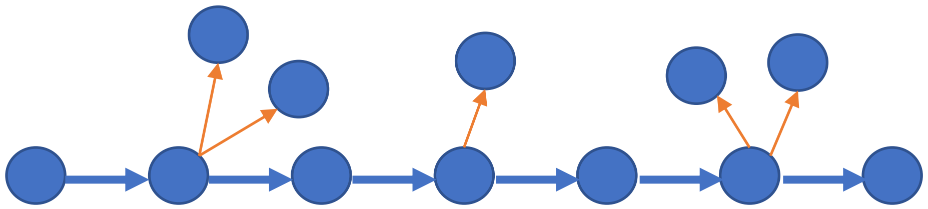
Diamond Attack:

1. Sep 3rd 1 address distributed ETH to 32 contributors
2. Donate *ProofOfStake* by Vitalik, GR15(6 grants)

Pipeline

1. Staring from the overall graph
2. Remove HUB addresses and smart contracts
3. Compute Weakly-Connect Components
4. For every WCC

5. Examine whether it contains Chain-Like attacks OR Diamond attacks



The heuristic to examine Chain is to prune branches so that the trunk is left as a Chain.

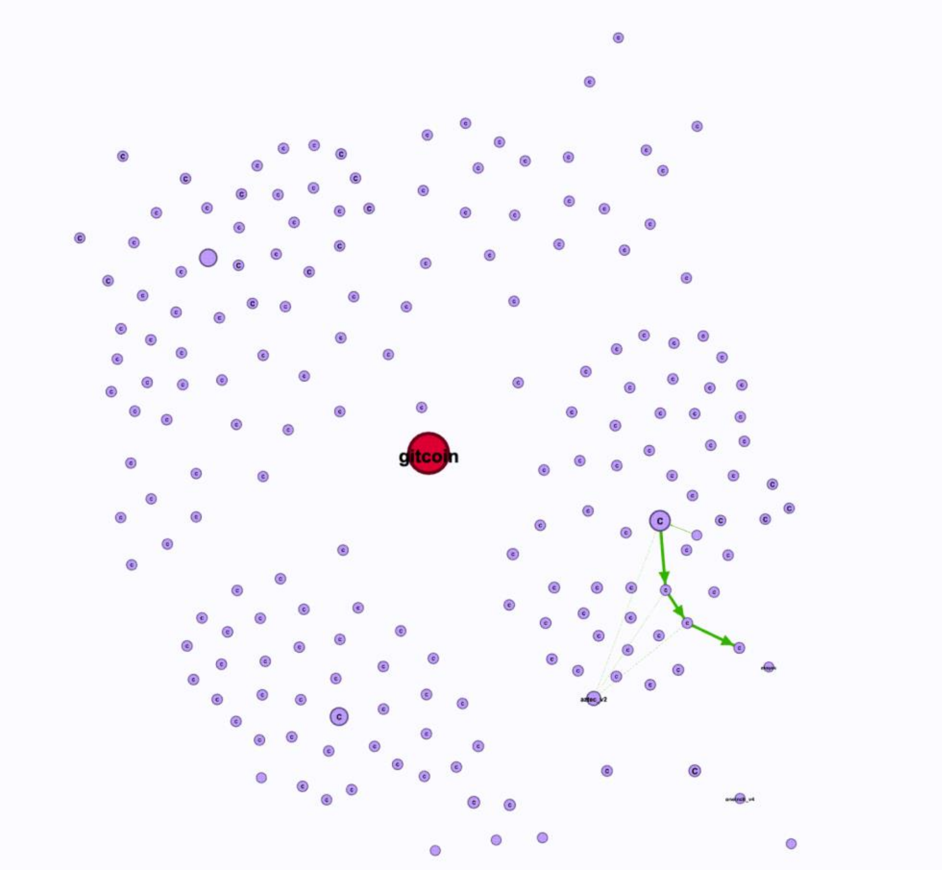
The key to find Diamond is to find the Fund Supplier with great out-degree and out-flow

6. Visualize possible Chain-Like attacks and Diamond attacks using Gephi
7. Manually double check them

Finally we have **150** Connect Components

Through carefully investigation, **2170** Addresses in the CCs are judged as Sybils **(Risk = HIGH)**

Example 3: Cluster With 188 Sybils



ClusterID=0 with 188 Addresses; Attack on Ethereum；

Fund Preparation:

Two addresses 0xede41(2022-09-14 9:25:49) and 0xb8308(2022-09-14 10:35:30) distributed fund in a Chain-like way.

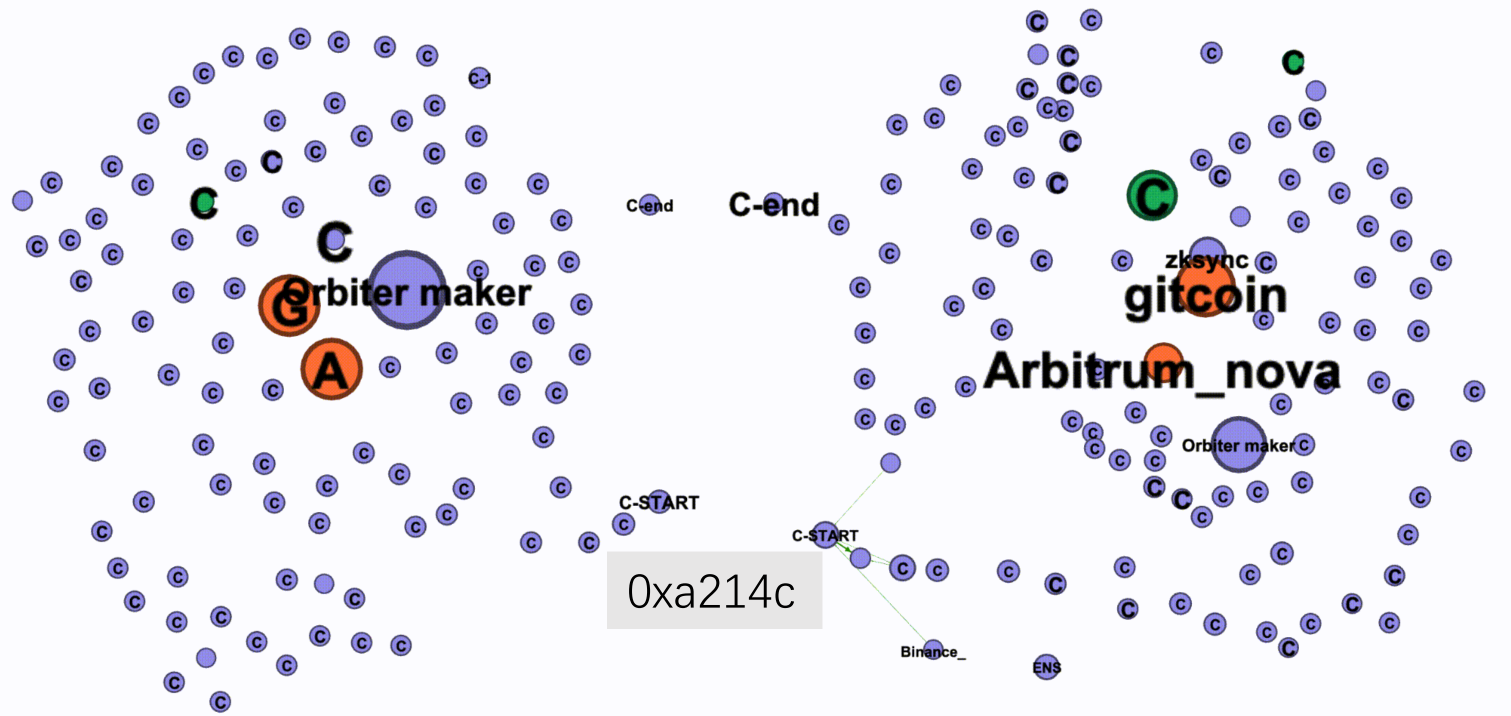
Making Donations:

With fund received, every address first made a donation of about 1USD, and then sent fund to the next.

Residual Collection:

After conducting attacks on Aztec and zkSync, all the residual funds are collected on sep 26th to three addresses on zkSync 0x45997, 0xede41 and 0xb8308.

Example 4: Cluster With 98 Sybils



Chain Attack on Ethereum

Chain Attack on zkSync

ClusterID=1； 98 Addresses；
On Ethereum and zkSync simultaneous

Fund Preparation:
Address C-START distributed funds in a Chain-like way, respectively on L1 Ethereum (09-15 16:16:23) and L2 zkSync (09-15 16:17:30)

Making Donations：
With fund received, addresses first made a donation to 2 grants, and then send funds to the next.

Residual Collection:
The last address along the chain transferred all residual funds back to C-START on Ethereum.

Agenda

1. Introduction
    1. Team
    2. Deliverables (github & demo video)
2. Our Works
    1. Data Preparation
    2. Topic 1：Bulk Transfers & Donations
    3. Topic 2： Sequential Behavior Pattern Mining
    4. Topic 3 ： Asset-Transfer Graph Mining
    5. Grant Fraud
3. Summary, Suggestion and Future Works

Grant Fraud

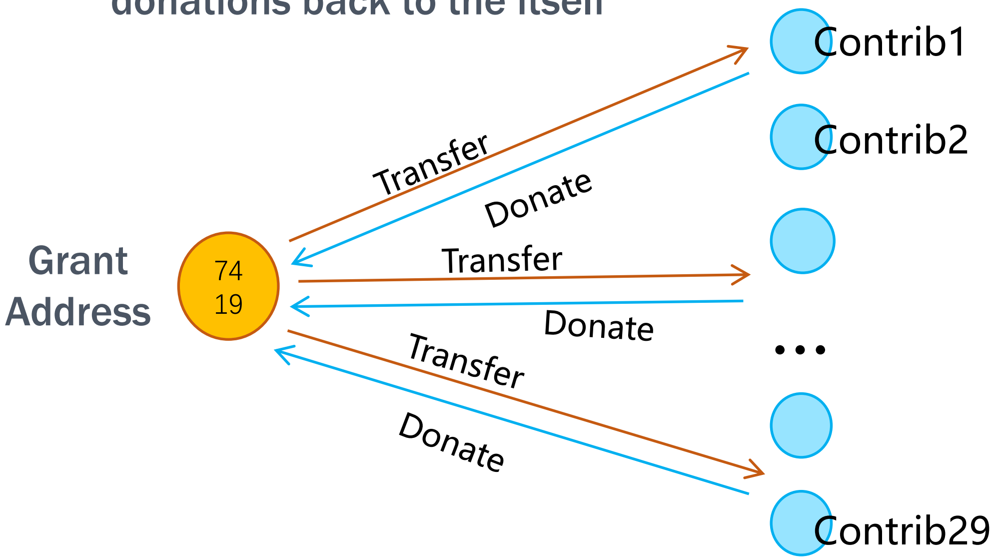
Basic Info

- Grant ID： 7419
- Contributors: 29
- Grant Address：

  0xfd9f8a0f4bdeac72f08af1c708023cc31dd2e3be

Reasoning

(1)The grant address transferred fund to 29 EOAs，and then (2) manipulated the EOAs to make donations back to the itself



Attack Details

Supply funds to EOAs (29 transfers)

0xd66710decabf00efe85...	Transfer	15516029	2022-09-11 17:15:20	48 days 11 hrs ago	0xfd9f8a0f4bdeac72f08a...	IN	0xcf1dc2cb1a5b5344330...	0.014 Ether
0x877427ce02e78a7218...	Transfer	15516034		48 days 11 hrs ago	0xfd9f8a0f4bdeac72f08a...	IN	0x0b6426070a98451308...	0.014 Ether
0x6cf20256188bffdf34e4...	Transfer	15516035	2022-09-11 17:17:11	48 days 11 hrs ago	0xfd9f8a0f4bdeac72f08a...	IN	0x288819c32f2228203a...	0.014 Ether
0x87e838f1249811a342...	Transfer	15516036	2022-09-11 17:18:01	48 days 11 hrs ago	0xfd9f8a0f4bdeac72f08a...	IN	0x05ed44153d4cb72748...	0.014 Ether
....								
0x6a86248fc1a6733788...	Transfer	15516042	2022-09-11 17:19:04	48 days 11 hrs ago	0xfd9f8a0f4bdeac72f08a...	IN	0x79f2d17c3f46a9fffbbd...	0.014 Ether

Bulk donations (39 donations)

address	time	chain	token	amount
0xfc66a1f969bb77eb89a314725d657312d58f1589	2022-09-10 23:29:03	eth_std	ETH	17.31
0x0b6426070a98451308ee54cd3b3c114f5d1a2d65	2022-09-11 17:28:42	eth_std	ETH	7.07
0x288819c32f2228203aa9065dfa53497cc2527e69	2022-09-12 03:12:57	eth_std	ETH	1.94
0xb869898cd011593d3d52037b743131924375e0ae	2022-09-12 20:20:00	eth_std	ETH	2.07
0x05ed44153d4cb72748595ef915118772cf189553	2022-09-13 14:04:37	eth_std	ETH	104.45
0x83ac2bb284930f4a9acfffb7cfb0dc0c92b5ab97	2022-09-14 11:31:10	eth_std	ETH	6.47
....				
0xae86d0ad922a0abe16878913a71cdcd018a50b96	2022-09-19 14:47:38	eth_std	ETH	1.98
0x1b5dff786eaccf5a41bf922e64313a0f4a60dab9	2022-09-19 14:47:42	eth_std	ETH	1.98
0xcf1dc2cb1a5b5344330b01a188d0cdc62773fe5e	2022-09-19 14:51:38	eth_std	ETH	1.98

Agenda

1. Introduction
   1. Team
   2. Deliverables (github & demo video)
2. Our Works
   1. Data Preparation
   2. Topic 1：Bulk Transfers & Donations
   3. Topic 2： Sequential Behavior Pattern Mining
   4. Topic 3 ： Asset-Transfer Graph Mining
   5. Grant Fraud
3. Summary, Suggestion and Future Works

Summary and Future Works

Risk Level and Address CNT

Risk Level	Final Result	Bulk Donation Risk	ATG Risk	Behavior Risk	Bulk Transfer Risk
high	16,994	14,005	2,170	1,669	705
medium	15,728	16,040	275	-	2,671
low	22,581	25,258	52,858	53,634	51,927
Total	**55,303**	**55,303**	**55,303**	**55,303**	**55,303**

Summary

We propose and develop four approaches, namely bulk transfers pattern mining, bulk donations pattern mining, asset-transfer graph mining, sequential behavior pattern mining for slaying sybil. These approaches form a systematic algorithmic LEGO and totally find **16,994** High Risk Sybils. Besides, we make our first attempt to detect grant fraud and find one case.

Future works

Due to time limit, some of the algorithms used in this work is not state-of-art. We can have more deep studies.

It would be very grateful if we can access Gitcoin exclusive data , e.g. ip, wifi, user behavior on Gitcoin etc.


HAPPY
HALLOWEEN!