black hat ASIA 2025

APRIL 3-4, 2025
BRIEFINGS

Sweeping the Blockchain: Unmasking Illicit Accounts in Web3 Scams

Speaker: Wenkai Li

Hainan University, China

Collaborators: Zhijie Liu (ShanghaiTech University),

Xiaoqi Li* (Hainan University)

*Corresponding author: csxqli@ieee.org

The Team

Li Wenkai

— PhD Student, Hainan University, China

— cswkli@hainanu.edu.cn

— https://cswkli.github.io/

Liu Zhijie

— Msc Student, ShanghaiTech University, China

— liuzhj2022@shanghaitech.edu.cn

— https://rroscha.github.io/

Li Xiaoqi

— Associate Professor, Hainan University, China

— csxqli@ieee.org

— https://csxqli.github.io/

About Us

Security Research

- 9 year-experience (since 2016) of Ethereum (Born in 2015) Blockchain Security.
- Blockchain/Software/System Security and Privacy, Ethereum/Smart Contract, Malware Detection, and etc.
- 40+ papers including ASE、INFOCOM、ICSE、WWW、AAAI、TSE, etc. within 5 years
- 30+ CVE/CNVD Vulnerabilities identified within 5 years
- 3700+ citations within 5 years
- Best Paper from INFOCOM、ISPEC、CCF, etc.
- SV Insight Annual Global Top-50 Blockchain Research Paper
- ESI Hot (Top 0.1%)、Highly Cited Paper (Top 1%)

Overview

Agenda

Introduction

Motivation

ScamSweeper

Experiments

Case Study

Introduction

The 3rd Generation Internet – Web 3.0

- Many ways for crypto users to engage with Web3.0:

NFT

Decentraland

Horizon Worlds

META

- The most used Web3.0 Services:

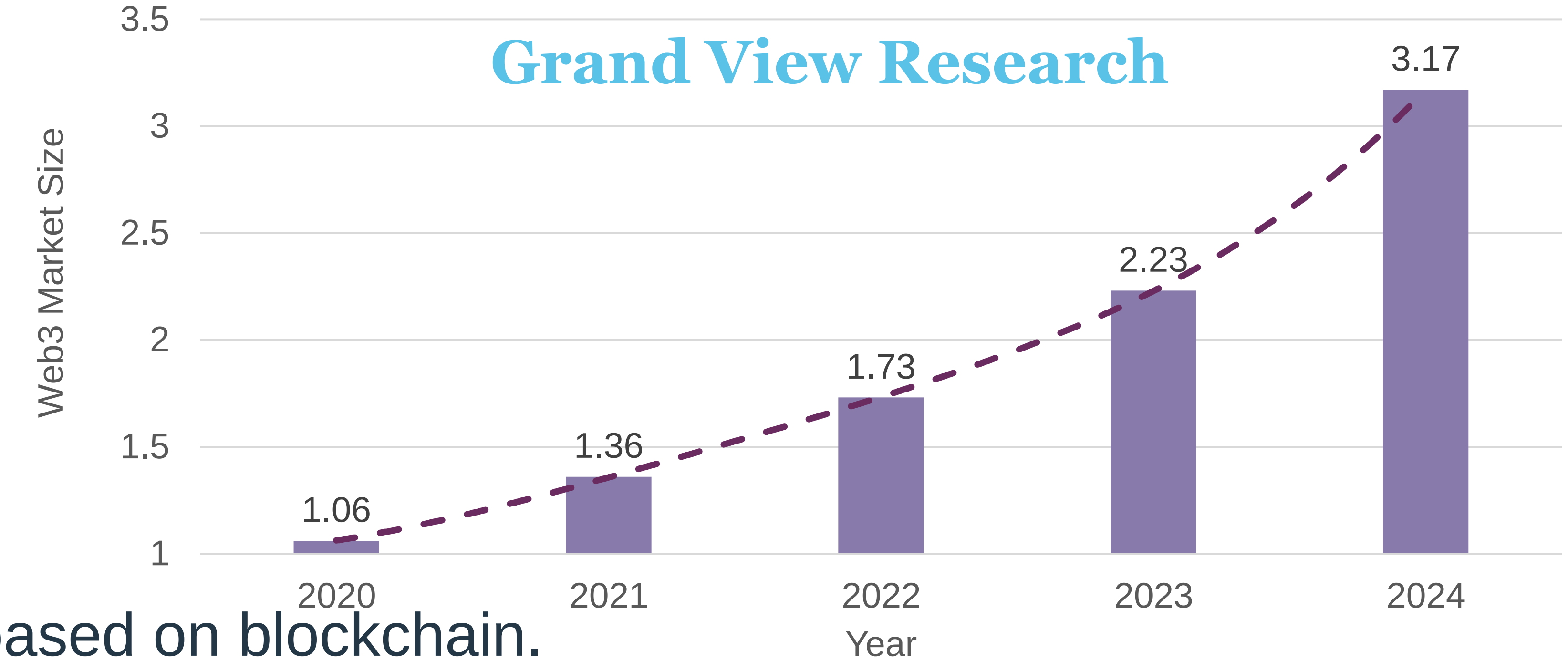
DEX

CEX

Crypto Gaming

The 3rd Generation Internet – Web 3.0

- What is the scale of Web3.0 tech market?
  - ➢ A growing trend.
  - ➢ The accelerating growth rate.
  - ➢ USD 3.17 billion in 2024.



- Web3.0 applications
  - ➢ DApp, DeFi protocol, DID, and etc. based on blockchain.
  - ➢ The blockchain node network follows a power-law distribution.
  - ➢ A minority of accounts appear at majority of Txs.

The Web3 environment comes with ***scam risks*** …

Motivation

Motivation: Web3 Scams

- The situation of Web3 scams:
  - ➢ Phishing, Rug Pulls, Harmful Airdrops, Giveaway Scams…
  - ➢ Crypto Drainer, Pig Butchering, Address Poisoning Scams…

NEWS 6 JAN 2025 www.infosecurity-magazine.com

Scammers Drain $500m from Crypto Wallets in a Year

- The scams on Web3 ecosystem can be catastrophic

Crypto Phishing Report 2024

Annual Report 2024 by ScamSniffer

$494M+ ↑67%
TOTAL LOSS

332K+ ↑3.7%
VICTIMS

$55.4M ↑130%
LARGEST SINGLE LOSS

30 ↑58%
LARGE LOSS CASES

ScamSniffer

NEWS 22 DEC 2023
Crypto Drainer Steals $59m Via Google and X Ads

NEWS 12 MAR 2024
Victims Lose $47m to Crypto Phishing Scams in February

NEWS 16 JAN 2024
Inferno Drainer Spoofs Over 100 Crypto Brands to Steal $80m+

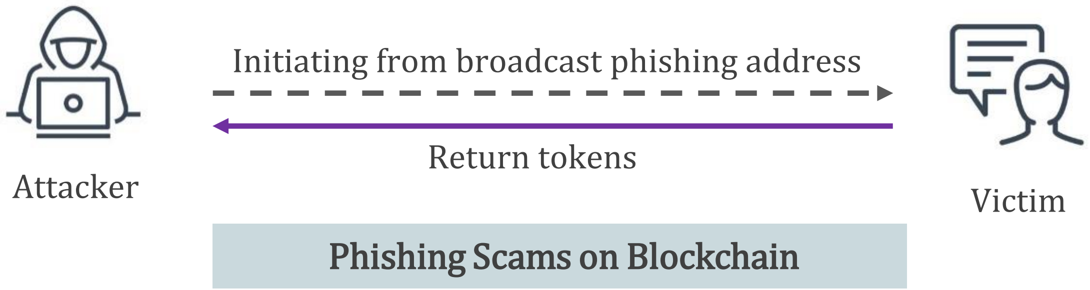
NEWS 8 JAN 2024
Security Firm Certik's Account Hijacked to Spread Crypto Drainer

NEWS 3 JAN 2025
Web3 Attacks Result in $2.3Bn in Cryptocurrency Losses

Motivation: Web3 Scams

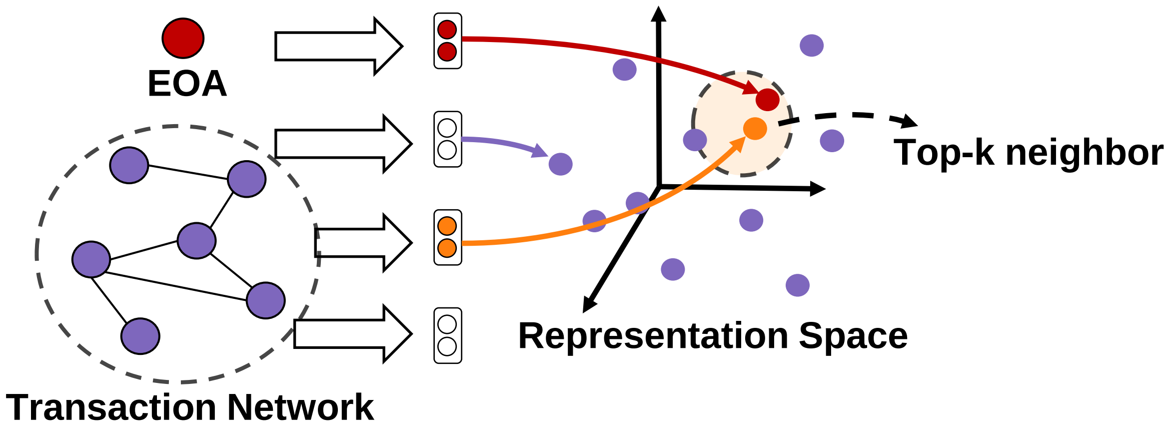
SCAM ALERT!

- What do the **Web3 Scams** on blockchain look like?
  - ➢ e.g., crypto drainers often masquerade as web3 projects, enticing victims into the drainer and getting the control access.

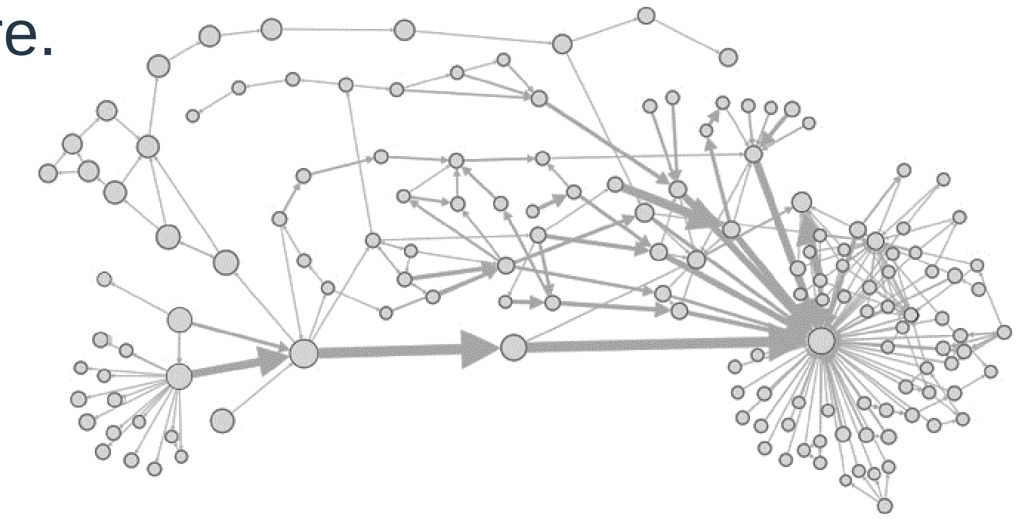


Motivation: Previous Research

- Graph Learning Methods
  - ➢ Intuitive to represent interactions of the topology structure.
  - ➢ Account as node, transaction as edge.
  - ➢ Top-k algorithm.
  - ➢ Power-law distribution leads lots of noise.



Top-k Algorithm

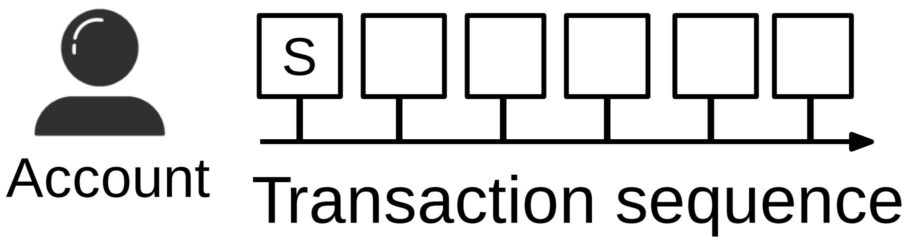


Random Walk

[1] Li, Shucheng and et al. "SIEGE: Self-Supervised Incremental Deep Graph Learning for Ethereum Phishing Scam Detection." in *Proc. of MM*. 2023.
[2] Wu, Zhiying and et al. "TRacer: Scalable graph-based transaction tracing for account-based blockchain trading systems. " *TIFS*. 2023.
[3] Li, Sijia and et al. "TTAGN: Temporal transaction aggregation graph network for Ethereum phishing scams detection." in *Proc. of WWW*. 2022.

Motivation: Previous Research

- Sequence Learning Methods
  - Transductive to learn the logic of account behavior feature.
  - Analyzing an account is related to its length.
  - Large-scale transactions, e.g., **2.7 billion txs** on Ethereum.



Tab.1 – The statistical information of some accounts on Ethereum.

No.	Account Address	Tx Cnt
1	0xC02aaA39b223FE8D0A0e5C4F27eAD9083C756Cc2	16,514,200
2	0x28C6c06298d514Db089934071355E5743bf21d60	18,921,592
3	0x267be1C1D684F78cb4F6a176C4911b741E4Ffdc0	3,832,284
4	0x32400084C286CF3E17e7B677ea9583e60a000324	3,094,481
5	0xf7858Da8a6617f7C6d0fF2bcAFDb6D2eeDF64840	1,588,678
6	0xA7EFAe728D2936e78BDA97dc267687568dD593f3	3,482,451
7	0xBf94F0AC752C739F623C463b5210a7fb2cbb420B	1,611,882
8	0xae0Ee0A63A2cE6BaeEFFE56e7714FB4EFE48D419	1,798,762
9	0x0D0707963952f2fBA59dD06f2b425ace40b492Fe	7,527,833
10	0x6262998Ced04146fA42253a5C0AF90CA02dfd2A3	1,183,120

[4] Hu, Sihao and et al. "BERT4ETH: A Pre-trained Transformer for Ethereum Fraud Detection." in *Proc. of WWW*. 2023.

Motivation: Previous Research

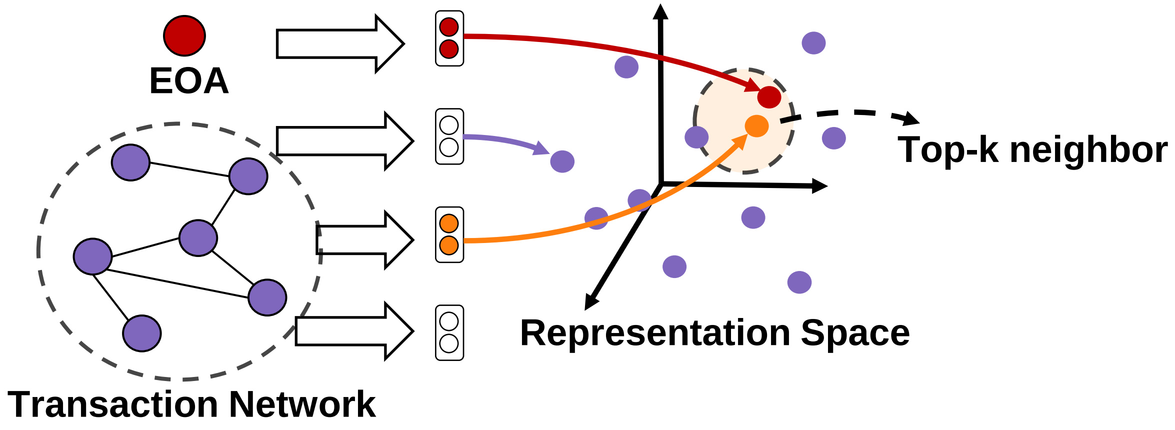
- Graph Learning Methods
  - ➢ Not suitable to capture dynamic information.

    Merging multiple edges into one for graph computation

    e.g., graph convolution or random walk
  - ➢ Not suitable for power law distribution.

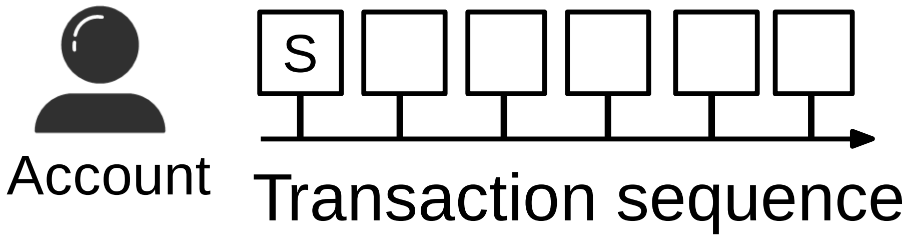
    Introducing noise when multi-hop convolution,

    In GRU, Model capability is limited (# of GNN layers = # of hop)



- Sequence Learning Methods
  - ➢ Not suitable to large-scale transactions.

    Analyzing an account is related to the length of its transaction sequence.

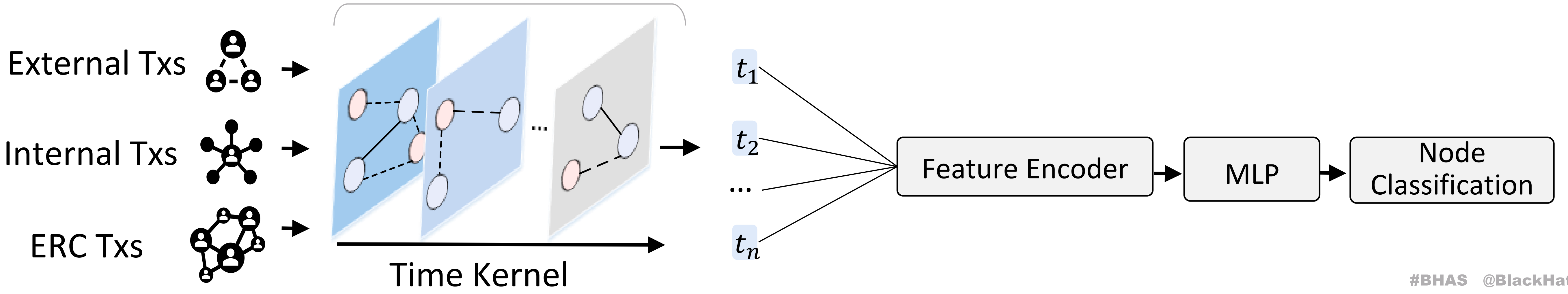


[4] Hu, Sihao and et al. "BERT4ETH: A Pre-trained Transformer for Ethereum Fraud Detection." in *Proc. of WWW*. 2023.

ScamSweeper

ScamSweeper

- Learning the dynamic evolution of transaction graph, and applying to account detection
  - ➢ Sequence learning from the graph structure.
  - ➢ crawl the data from the Etherscan, Ethereum and github.
  - ➢ transaction network construction.
  - ➢ split the graph into several subgraph according to the temporal series.
  - ➢ learn each sub-graph feature.
  - ➢ capture the dynamic evolution, and make a classification.



ScamSweeper (1)

- (a) Graph Construction
  - Most previous works used the **random walk** to sample the transaction network.
  - Random walk is like a ***dice game!***

Motivation:

To lower the computing consumption, and learn features from temporal sequence and topology structure.

We designed a new walk-sampling method:

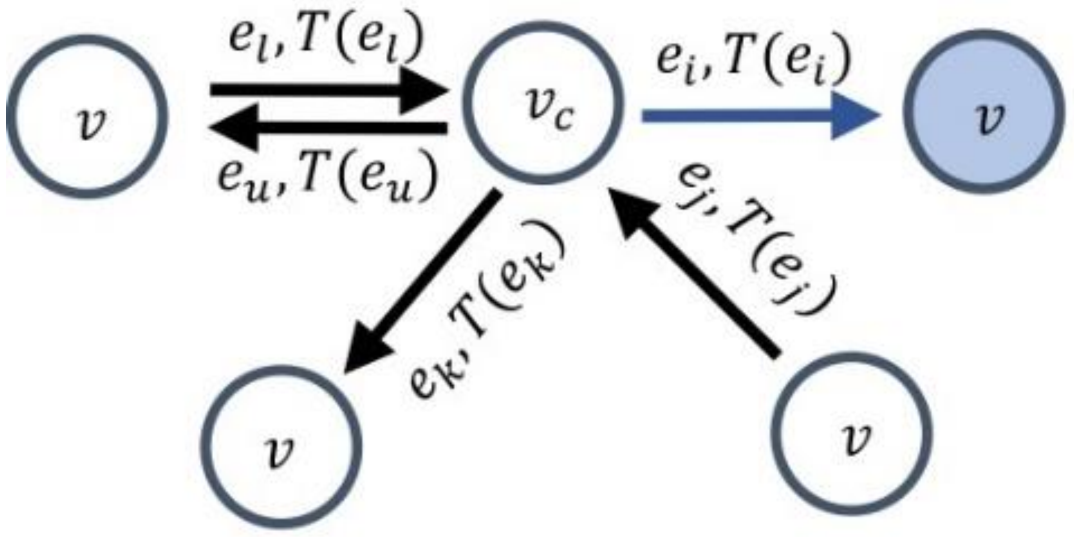
Struct-Temporal Random walk (STRWalk)

ScamSweeper (1)

- (a) Graph Construction
  - current node is $\boldsymbol{v_i}$, next node is $\boldsymbol{v_{i+1}}$,
  - the edge is $\boldsymbol{e_i}$
  - $\mu(T(\boldsymbol{e_i}))=T(\boldsymbol{e_i})-$*mintime*,
  - $\boldsymbol{\delta}(\boldsymbol{v})$ represents the number of nodes that are in the same interval with $\boldsymbol{v}$.

With $\boldsymbol{P_i}$ and $\boldsymbol{p_m}$, Struct-Temporal Random walk (STRWalk)

With $\boldsymbol{P_i}$, Temporal Random Walk (TRWalk)



The 1st sampled node selected by the alias sample algorithm with the **probability $\boldsymbol{p_i}$**.

The 2nd sampled node selected by the alias sample algorithm with the **probability $\boldsymbol{p_m}$**.

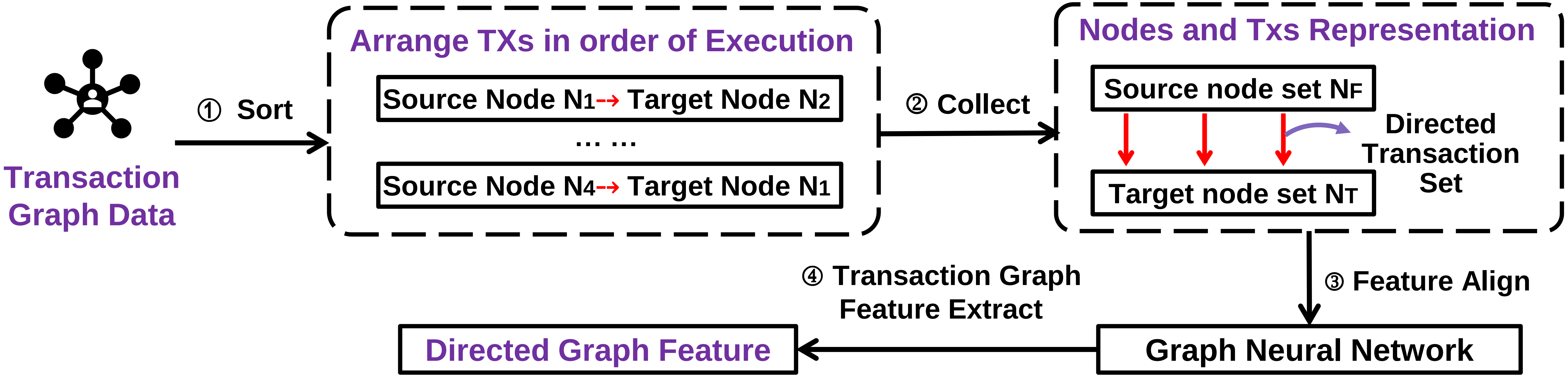
ScamSweeper (1)

- (a) Graph Construction
  - ➢ Walk length: 20, the window size: 4, and the embedding dimension: 128
  - ➢ Phishing dataset, 1165 **malicious** nodes and 636 **normal** nodes.
  - ➢ T-SNE Visualization

- Results
  - ➢ Random walk and deep walk are selected into ScamSweeper at the same time as TRWalk and STRWalk
  - ➢ The model with STRWalk can segment almost linearly.

ScamSweeper (2)

- (b) Directed Graph Encoder
  - Spliting the whole graph according to the interval, generating several sub-graphs
  - Learning the feature of each subgraph in time sequence



ScamSweeper (3)

- (c) Temporal Feature Learning
  - Leveraging the ability of Transformer

$$H^{(l+1)} = Attention({H^{(l)}}^T \Theta_Q, {H^{(l)}}^T \Theta_K, {H^{(l)}}^T \Theta_V) \quad (5)$$

$$h = Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}} V) \quad (6)$$

$$H^{(l+1)} = FFN(h) \quad (7)$$

$$FFN(x) = Sigmiod\left(xW_1^{(l)} + b_1^{(l)}\right) W_2^{(l)} + b_2^{(l)} \quad (8)$$

  - sort them in the original time order.
  - variational Transformer structure to extract the features on time.
  - capture the dynamic evolution feature.

Experiments

Experiments: Large-scale Data

- Data & distribution
  - Crawling the first 18 million block height on Ethereum
  - Phishing labels from Etherscan
  - Web3 scams from [5]
  - Normal nodes contains 4 types: exchange, mining, ICO wallet, and gambling.

[5] https://github.com/scamsniffer/scam-database, 2024.

Experiments: Ablation

- How well do the components work?
  - the importance of graph encoder and T-Transformer
  - The ScamSweeper is mainly contained with both of them.

- Results
  - Under the F1-score and Weighted F1-score
  - Both of them are valuable for the classification.

ScamSweeper > Graph encoder > T-Transformer

Experiments: Comparison

- How well do the ScamSweeper work?
  - Compared with Graph methods and Transformer
  - Structure window: {5,10,15}.
  - Training: 70%, Validation: 20%, Test:10%
  - STRWalk samples the network and then used for the evaluation

- Results
  - Under the Accuracy, F1-score, Precision, and Recall
  - scamSweeper always outperform other methods.

Case Study

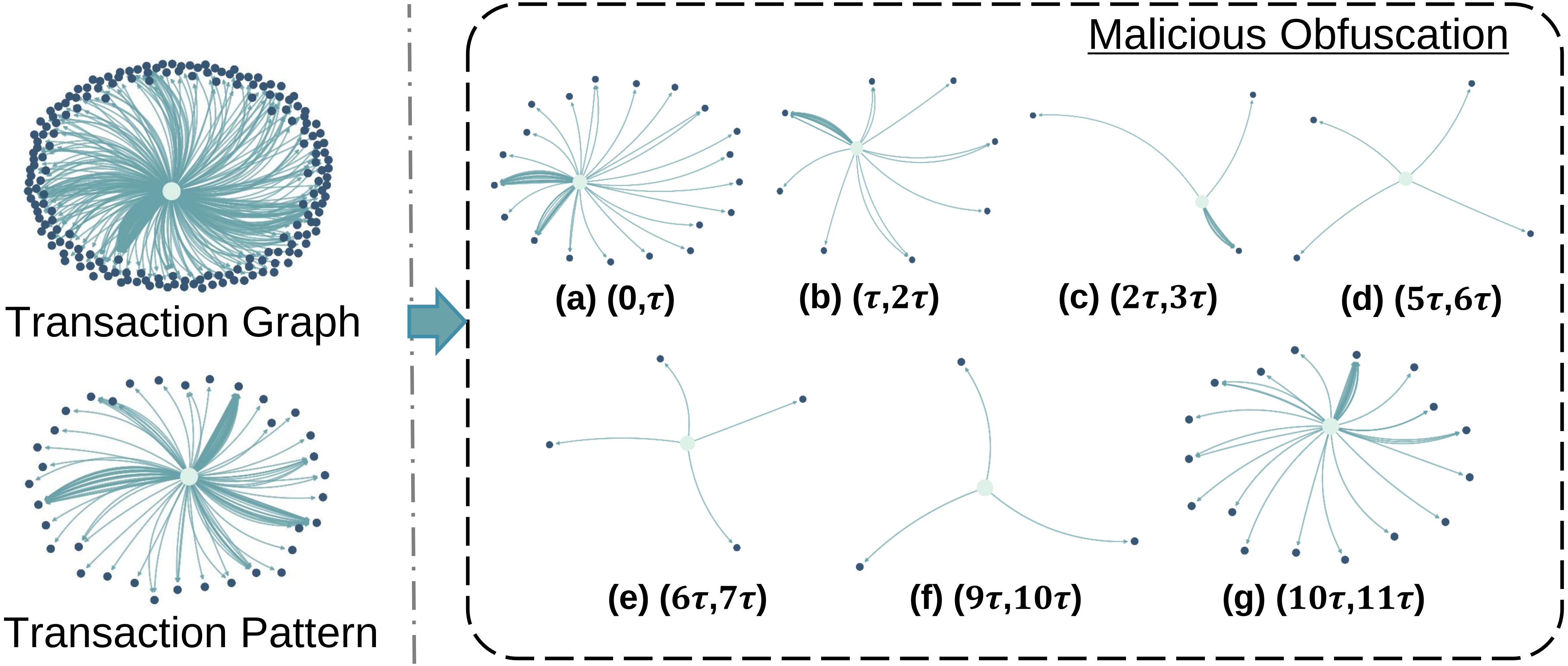
Case Study: Web3 Scam

- Dynamic Evolution
  - ➢ τ is a time interval.
  - ➢ The time interval sets 1 day.
  - ➢ All the data from on-chain data,
  - ➢ not contained any off-chain data.
  - ➢ scammer mimics the normal regularity
  - ➢ the decrease of transaction volume
  - ➢ Not suffer any DoS attack

	A	B	C	D
1	from	to	value	timestamps
2	0x075d9bl	0x375abb8	0	1.68E+09
3	0x075d9bl	0xf17a3fe5	0	1.68E+09
4	0x075d9bl	0x29488e5	2.27818	1.68E+09
5	0x075d9bl	0x29488e5	1.54884	1.68E+09
6	0x075d9bl	0x7749afe	0.010056	1.68E+09
7	0x075d9bl	0xc183602	0	1.68E+09
8	0x075d9bl	0x881d402	0	1.68E+09
9	0x075d9bl	0x881d402	0	1.68E+09
10	0x075d9bl	0x881d402	0	1.68E+09
11	0x075d9bl	0x881d402	0	1.68E+09
12	0x075d9bl	0x881d402	0	1.68E+09
13	0x075d9bl	0x881d402	0	1.68E+09
14	0x075d9bl	0x881d402	0	1.68E+09
15	0x075d9bl	0x881d402	0	1.68E+09
16	0x075d9bl	0x881d402	0	1.68E+09
17	0x075d9bl	0x881d402	0	1.68E+09
18	0x075d9bl	0x881d402	0	1.68E+09
19	0x075d9bl	0x881d402	0	1.68E+09
20	0x075d9bl	0x881d402	0	1.68E+09
21	0x075d9bl	0x881d402	0	1.68E+09
22	0x075d9bl	0x881d402	0	1.68E+09
23	0x075d9bl	0x881d402	0	1.68E+09
24	0x075d9bl	0x881d402	0	1.68E+09
25	0x075d9bl	0x881d402	0	1.68E+09
26	0x075d9bl	0x881d402	0	1.68E+09
27	0x075d9bl	0xd417144	0	1.68E+09
28	0x075d9bl	0xa0b8699	0	1.68E+09
29	0x075d9bl	0xa0b8699	0	1.68E+09
30	0x075d9bl	0x5668145	0	1.68E+09
31	0x075d9bl	0x0000000	0	1.68E+09
32	0x075d9bl	0x0000000	0	1.68E+09
33	0x075d9bl	0x10a5703	11.30202	1.68E+09
34	0x075d9bl	0x10a5703	7	1.68E+09
35	0x075d9bl	0xb1ca0f7	0.133012	1.68E+09
36	0x075d9bl	0x7d1afa7	0	1.68E+09
37	0x075d9bl	0xde30da3	0	1.68E+09
38	0x075d9bl	0x5149107	0	1.68E+09
39	0x075d9bl	0x1f9840a	0	1.68E+09
40	0x075d9bl	0x95ad61b	0	1.68E+09
41	0x075d9bl	0x03be5c9	0	1.68E+09
42	0x075d9bl	0xef1c6e6	0	1.68E+09
43	0x075d9bl	0xef1c6e6	0	1.68E+09
44	0x075d9bl	0xef1c6e6	0	1.68E+09
45	0x075d9bl	0xef1c6e6	0	1.68E+09
46	0x075d9bl	0xef1c6e6	0	1.68E+09
47	0x075d9bl	0xef1c6e6	0	1.68E+09
48	0x075d9bl	0xef1c6e6	0	1.68E+09
49	0x075d9bl	0xef1c6e6	0	1.68E+09
50	0x075d9bl	0xef1c6e6	0	1.68E+09
51	0x075d9bl	0xef1c6e6	0	1.68E+09
52	0x075d9bl	0xef1c6e6	0	1.68E+09

Case Study: Web3 Scam

- Dynamic Evolution
  - τ is a time interval.



Takeaways

- Summary & key takeaways

  - **Web3 Scams Proliferation:** Web3 applications are increasingly targeted by scammers who mimic legitimate transactions to deceive users, highlighting a critical gap in current detection methods.

  - **Research Gap:** Prior studies focus on de-anonymization and phishing nodes, neglecting the unique temporal and structural patterns of web3 scams, while existing detection tools struggle with power-law distributed transaction networks.

  - **New Approach:** A novel approach that combines structure-temporal random walks for efficient transaction network sampling and variational transformers for dynamic pattern analysis, capturing both temporal and structural evolution of scams.

  - **Practical Insights:** Large-scale dataset collection, effective data sampling, and dynamic evolution analysis, enabling real-world application in Ethereum transaction monitoring.

Thank You