SMU
SINGAPORE MANAGEMENT UNIVERSITY

School of
Computing and
Information Systems

Information and Knowledge Discovery on Cryptocurrency: A Security Perspective

Hands-on Tutorial at
The 2024 IEEE International Conference on Big data (IEEE BigData 2024)

Feida ZHU, PhD
Associate Professor

Ling CHENG, PhD
PhD Student

School of Computing and Information Systems
Singapore Management University

IEEE BIG DATA 2024
December 15-18 • Washington DC, USA

Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path

Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path

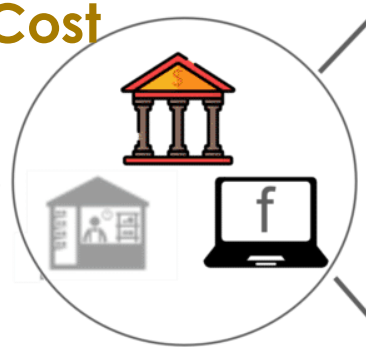
Why Blockchain?

Before Blockchain

1. High Trust Cost

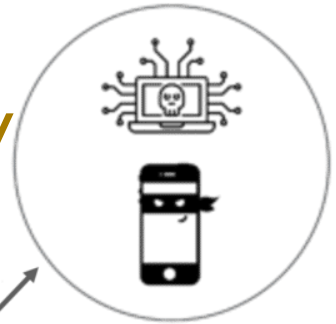
Sensitive Data- personal, financial, identity

Data given to firms online to complete transactions, create accounts etc



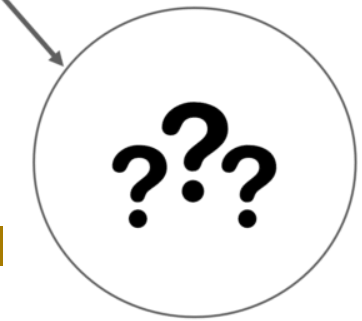
Online merchants, platforms, banks trusted with storing our sensitive data

3. High Vulnerability



Single entity 3rd party data storage is vulnerable to online hacking, leaks and fraud

2. Limited Control



Firms sell our sensitive data to unknown 3rd parties

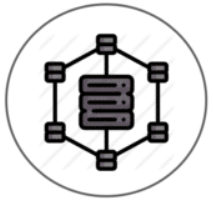
After Blockchain

Sensitive Data- personal, financial, identity

Sensitive data is encrypted on a distributed network instead of stored in a single instance database

You alone have the unique key to access your own data

You control which party should access your data



No single entity stores your data online

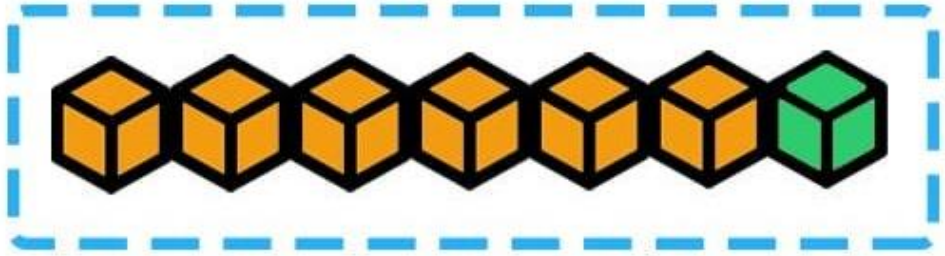
What is Blockchain?

Someone requests a **transaction**.

The requested transaction is broadcast to a **P2P** network consisting of computers known as **nodes**.

The P2P network of nodes validates the transaction and the user's status using **known algorithms**.

The transaction is complete!



The new block is then added to the existing blockchain in a way that is **permanent** and **unalterable**.

Once verified, the transaction is combined with other transactions to create a new **block of data** for the **ledger**.

A **verified** transaction can involve **cryptocurrency**, contracts, records, or other information.

Cryptocurrency

Has no **intrinsic value** in that it is not redeemable for another commodity.

Has no physical form and exists **only in the network.**

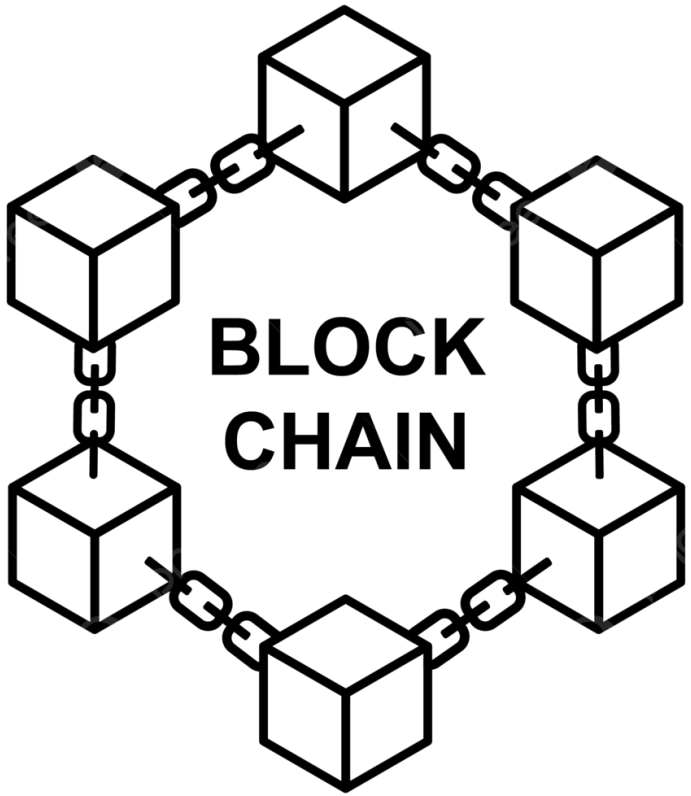
Its supply is not determined by a central bank, and the network is **completely decentralized.**

1. https://medium.com/@rajasinghs/what-is-blockchain-technology-a-step-by-step-guide-for-beginners-1393c992fc1b

Blockchain Properties

Programmable

A blockchain is programmable. (i.e. Smart Contracts)



Distributed

All network participants have a copy of the ledger for complete transparency.

Unanimous

All network participants agree to the validity of each of the records

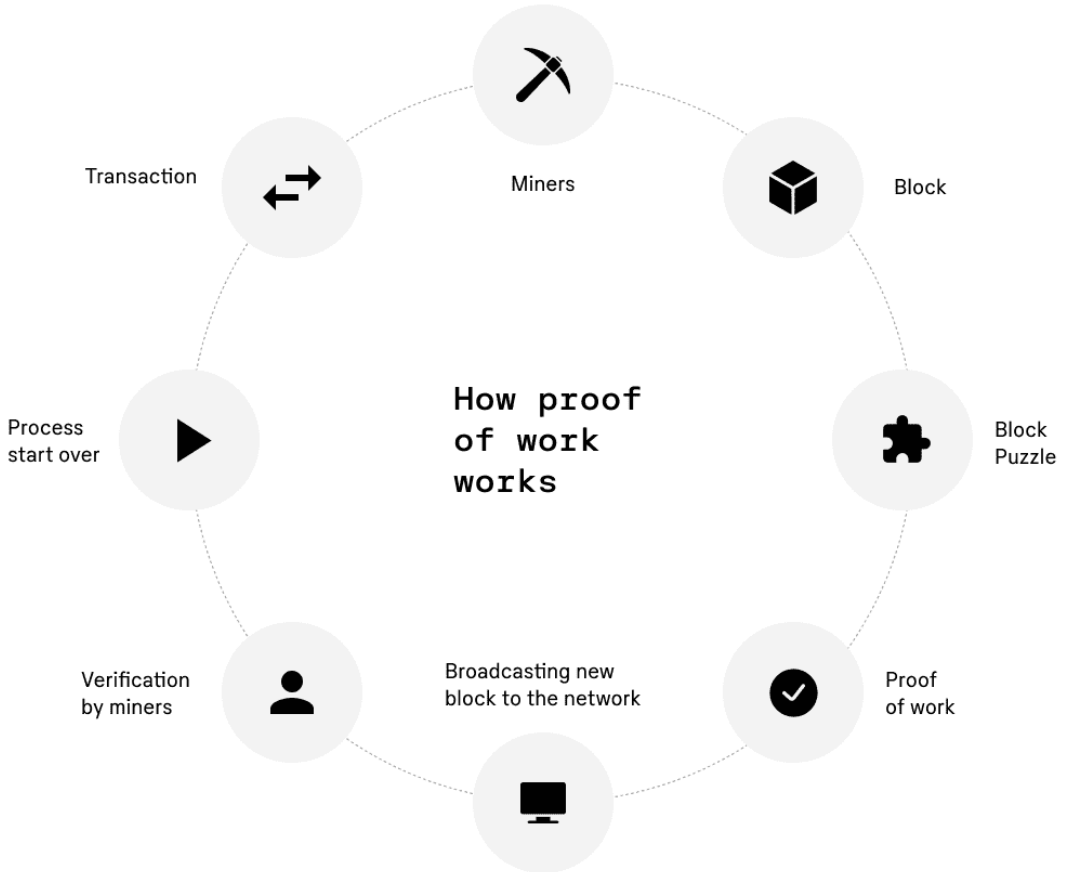
Immutable

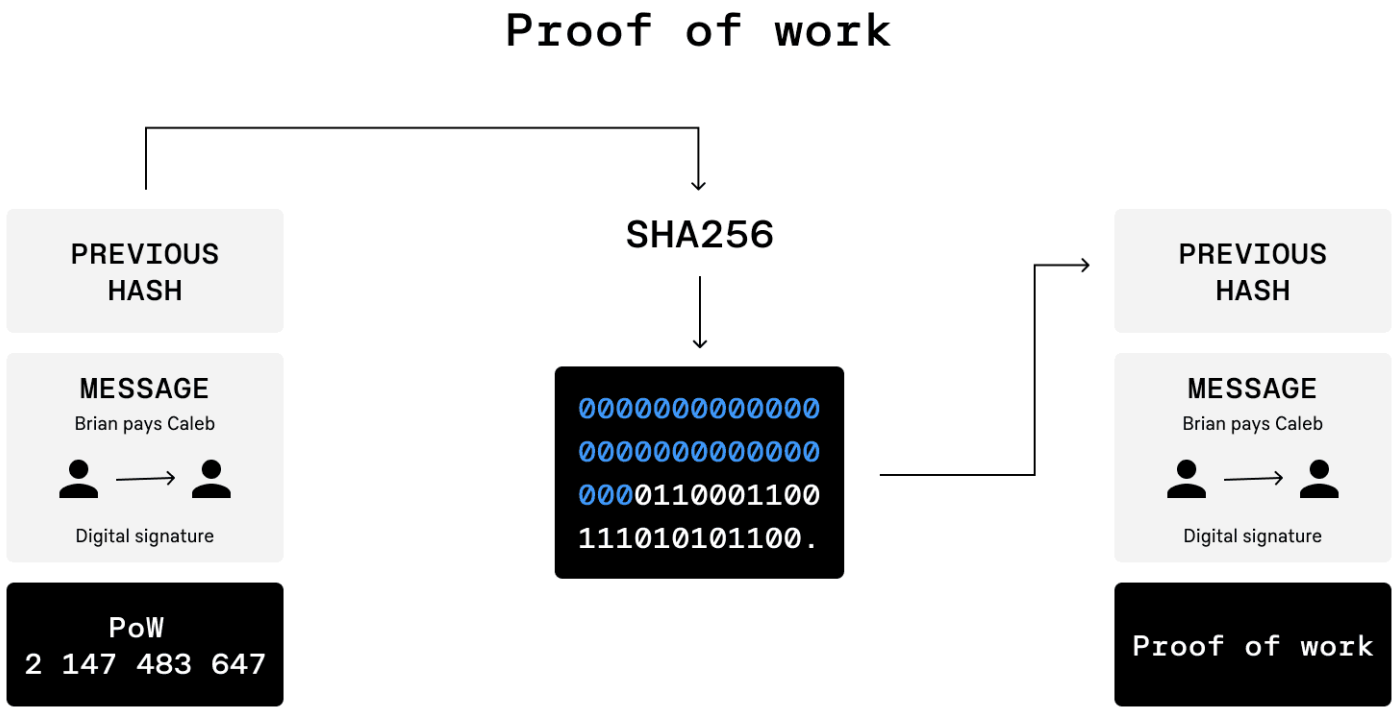
Any validated records are irreversible and cannot be changed.

Anonymous

The identity of participants is anonymous

What is Proof of Work?

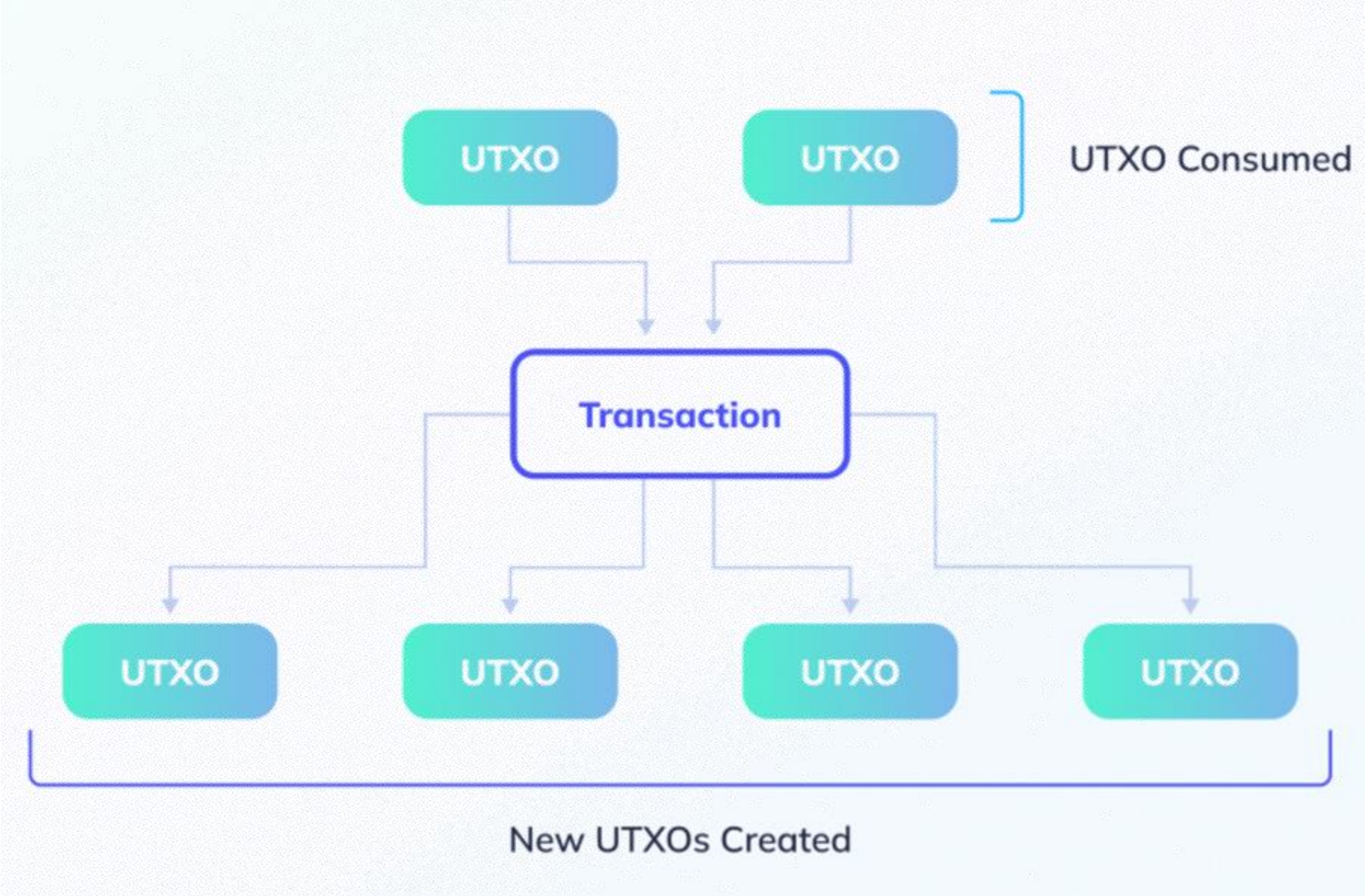




Online Demo Link

1. https://coinloan.io/blog/what-is-proof-of-work-pow/

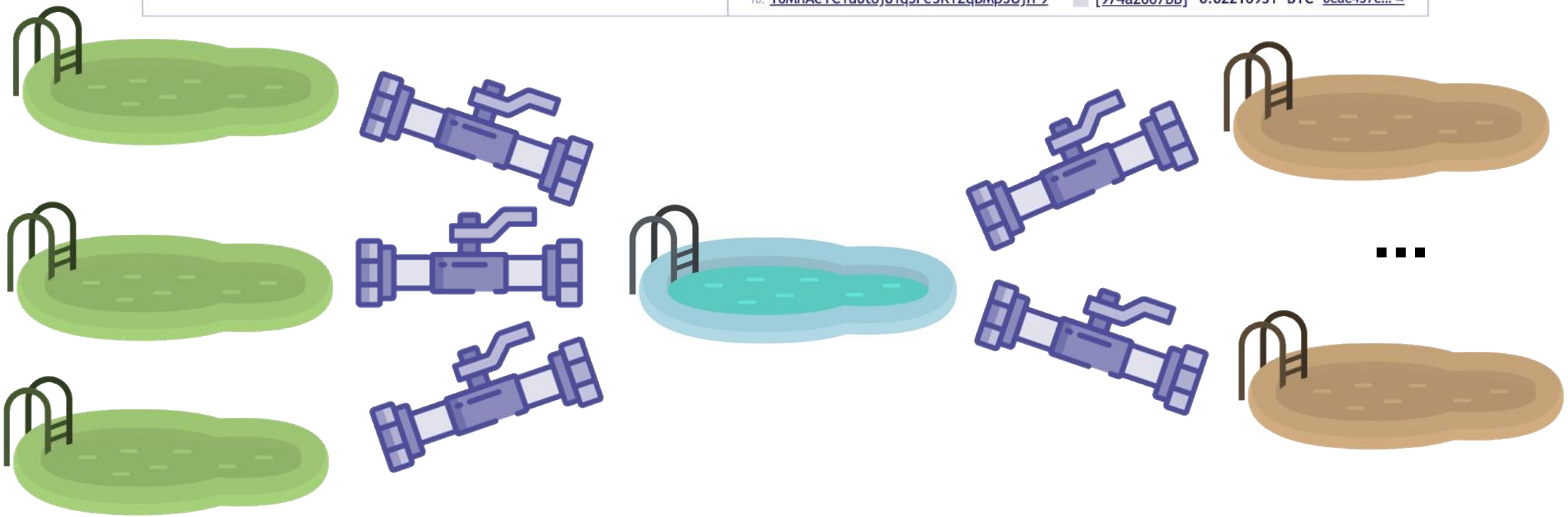
What is Proof of Stake?



BTC UTXO Transaction System

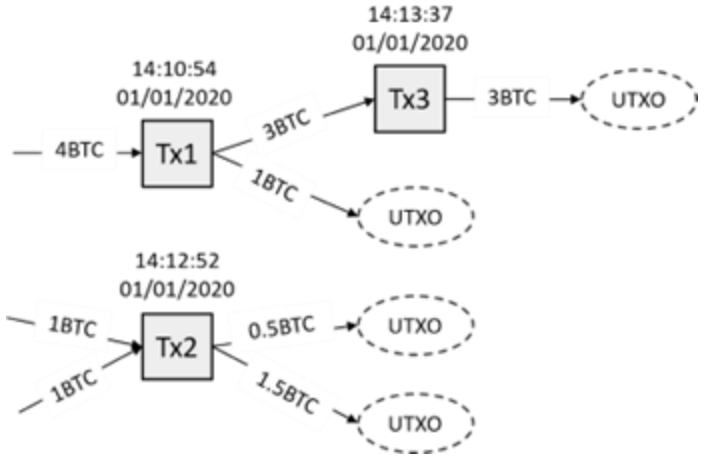
inputs: 3 (0.23154944 BTC)	unique addresses: 3, source transactions: 3		
0. 1AJCCxhKquHMdPW1Lw3RdkqXj9JVbGdQyg	0.16311193 BTC	⇐ 99411a66...	
1. 1CwXs1tbv5Dv3DN1HWLii7thnjJ6DaVXfv	0.01012727 BTC	⇐ e224de23...	
2. 1QGK4uWK43wWkheKyjMJHkUHWnmK8X2J6L	0.05831024 BTC	⇐ 4fff7a1c...	

outputs: 11 (0.23072744 BTC)	unique addresses: 11, spent: 11 in 10 transactions		
0. 37CUkHnZL1fana8DVUBFTb9fj5kmk5amv9	Xapo.com	0.02189381 BTC	b917046e... ⇒
1. 1CDMadTT4mzt2yrEBSKi78XFkdnfaiqy4h	BitoEX.com	0.02212267 BTC	49fdc519... ⇒
2. 3PiAFWjTj8C29PbaKDju6SVb342rZZy1wd	[000004b74b]	0.02232449 BTC	0c13c42f... ⇒
3. 15EqXrJzeFhw63RBACXD5s4DgfcPvCpxvQ	[0004fdeba6]	0.02212267 BTC	2316ed2c... ⇒
4. 12YKAG1ARHFJTBnHu3v3pWgg3dZXMup2i8	(change address)	0.02189381 BTC	47ddc53b... ⇒
5. 12Wsn7Dvi8r97uDRyRaMptkQKCvh8mD591	BitBay.net	0.02211753 BTC	7140669f... ⇒
6. 1P92nSacpTn9VaVTvTkR5akbV5dpeuZU8R	[12ce0a21e7]	0.02197031 BTC	8b51d3f6... ⇒
7. 3JLEyDXPV14oW8N526PLX6tAAcRFde9Cp1	[f041f40b3e]	0.02211496 BTC	a214787d... ⇒
8. 1AQdmPW9GDm5kn576RA4ifYPhFKNjmZnTL	BitBay.net	0.0219782 BTC	185df6c8... ⇒
9. 1NUxerSwgaAZ4LEx2YuT9B7rVFeHcDhDMD	(change address)	0.01001968 BTC	47ddc53b... ⇒
10. 16MhAeYCYu8toj8YqSPe5K1ZqBMp3UjfF9	[974a2667bb]	0.02216931 BTC	0cac457c... ⇒

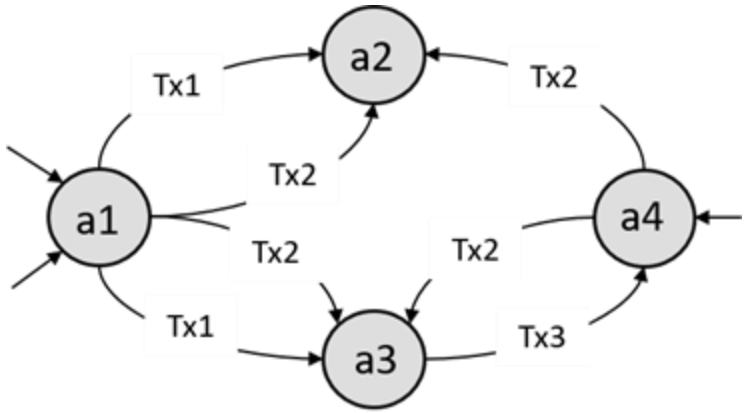


BTC UTXO Transaction System

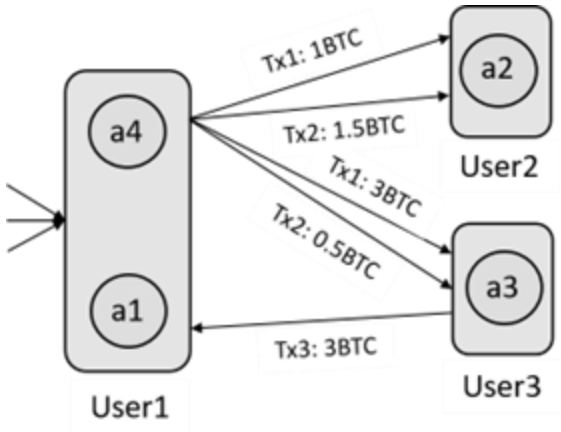
➢ BTC Transaction Network



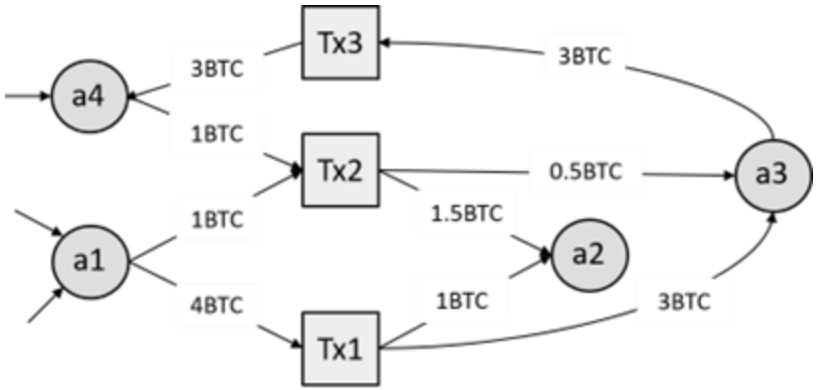
Transaction Network



Address Network



User Network



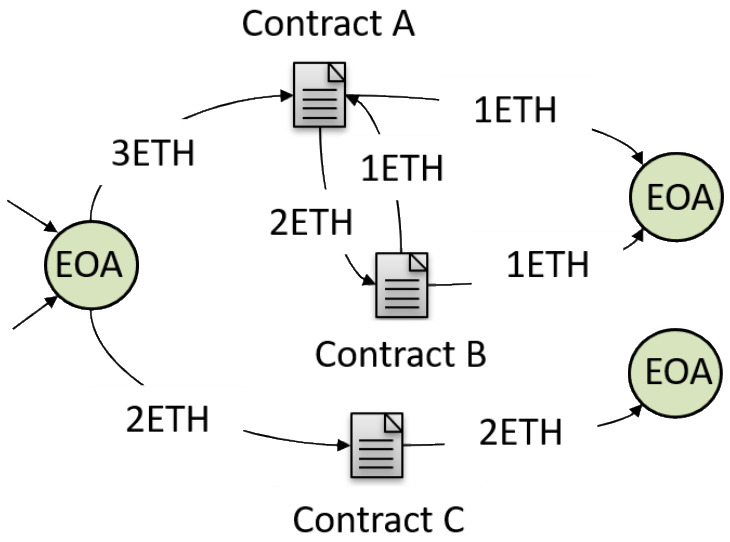
Hypergraph

1. Wu, Reid, F., Harrigan, M. *"An analysis of anonymity in the Bitcoin system"*. Springer. pp. 197–223 (2013).

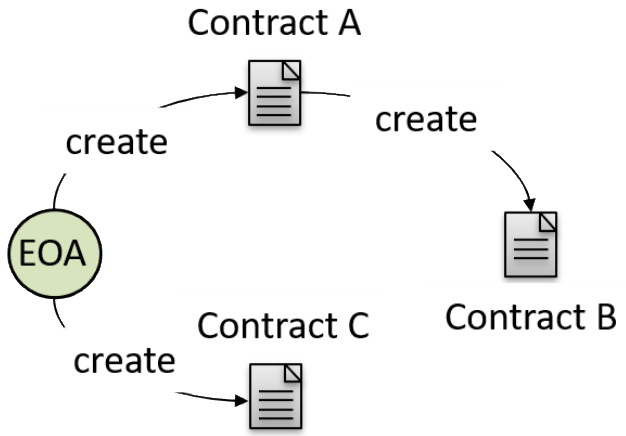
2. Pham, T., Lee, S. *"Anomaly detection in the Bitcoin system-A network perspective"*. arXiv preprint (2016).

ETH Account-based Transaction System

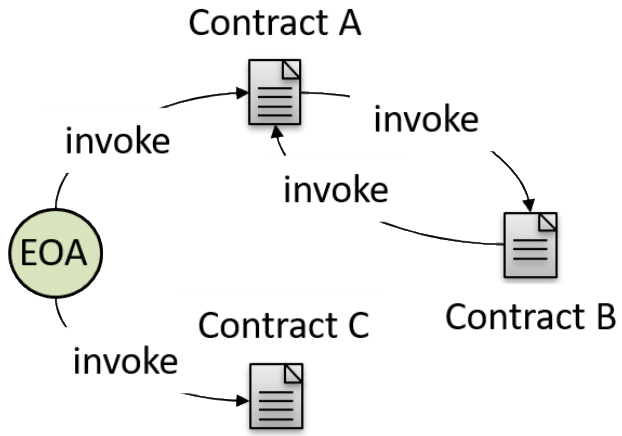
➢ ETH Transaction Network



Money Flow Graph



Contract Creation Graph

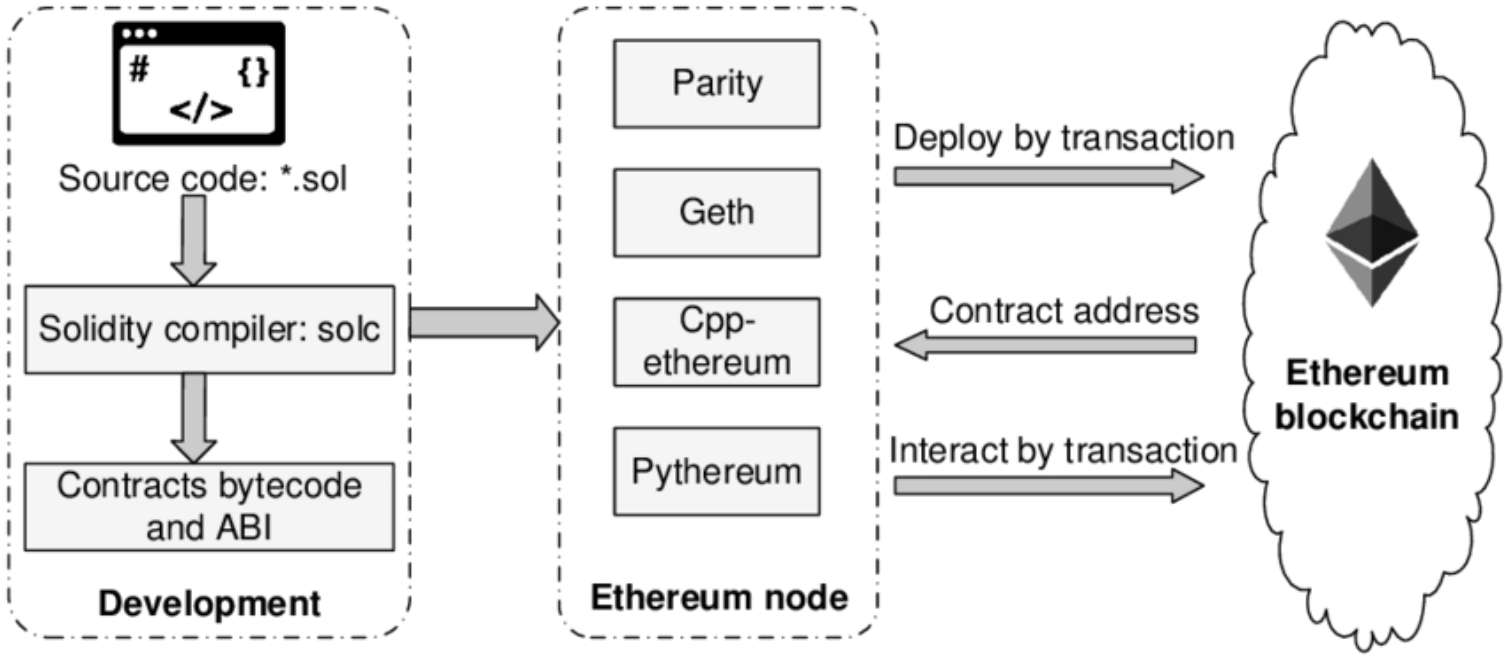


Contract Invocation Graph

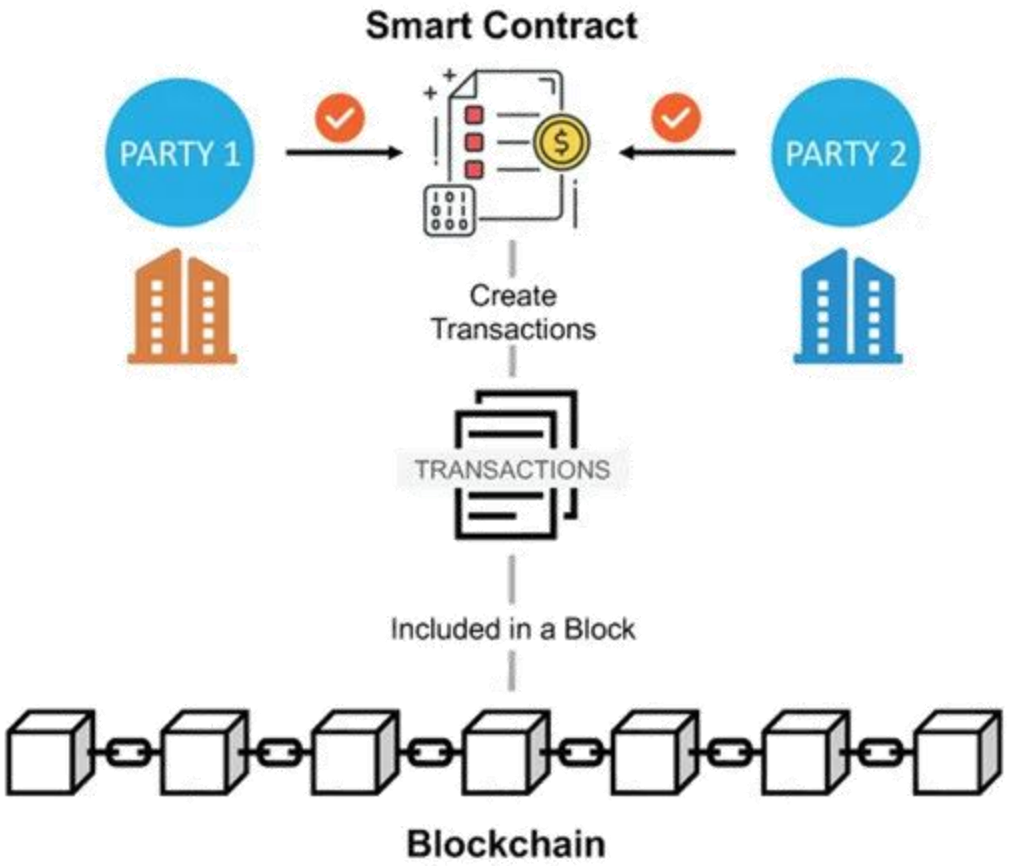
1. Chen, T. et al. *" Understanding Ethereum via graph analysis",* in: Proc. IEEE Conf. Comput. Commun(2018).

Smart Contract

➢ ETH Smart Contract Deployment



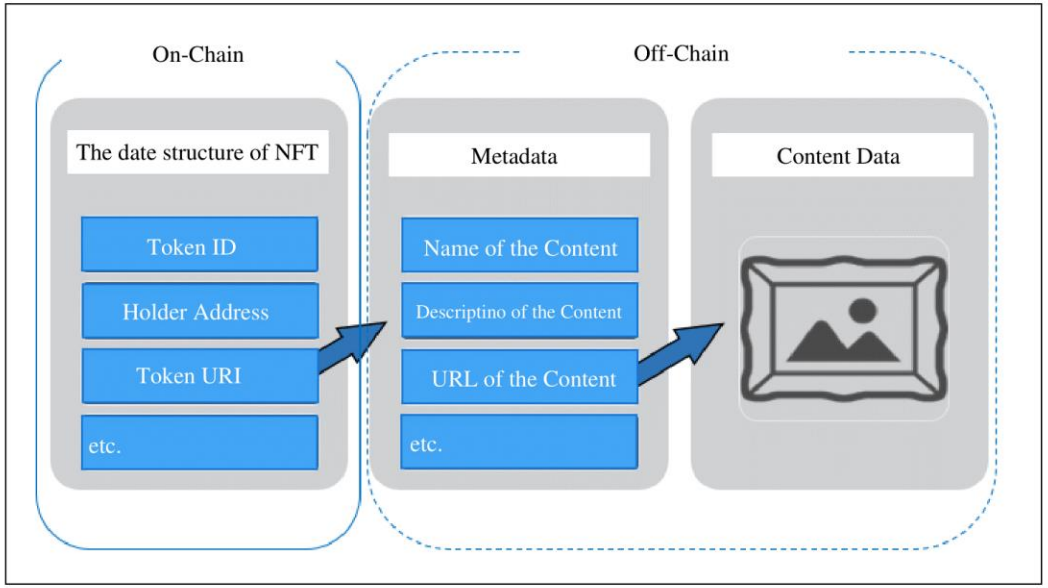
➢ Contract Interaction



Ethereum nodes are computers running Ethereum client software that participate in the Ethereum blockchain network. These nodes are responsible for **maintaining the blockchain**, **validating transactions**, **executing smart contracts**, and **sharing data** across the network.

Non-Fungible Tokens (NFT)

- A smart contract (e.g., ERC 721) to bond together the following information
  - A unique identifier (identity)
  - A set of standard metadata to display the asset
    - Name
    - Description
    - Image
  - A secure file link (content)

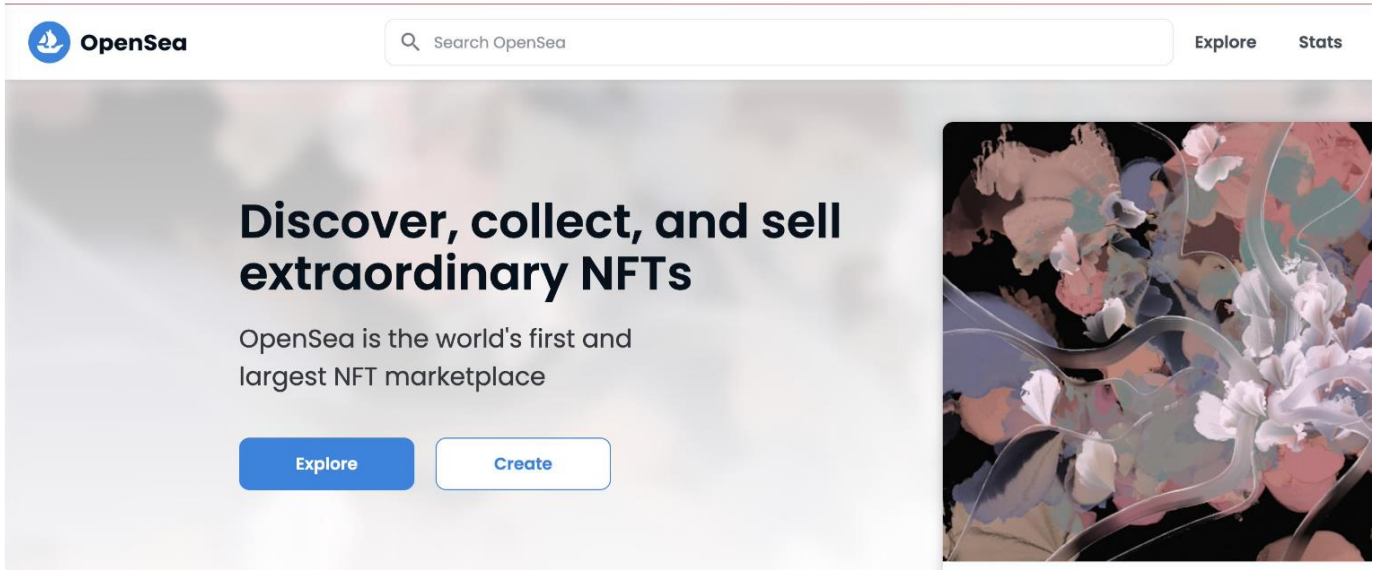


- Metadata and associated files are deployed onto the InterPlanetary File System (IPFS), a secure, transparent, decentralized, and public blockchain to host asset metadata.

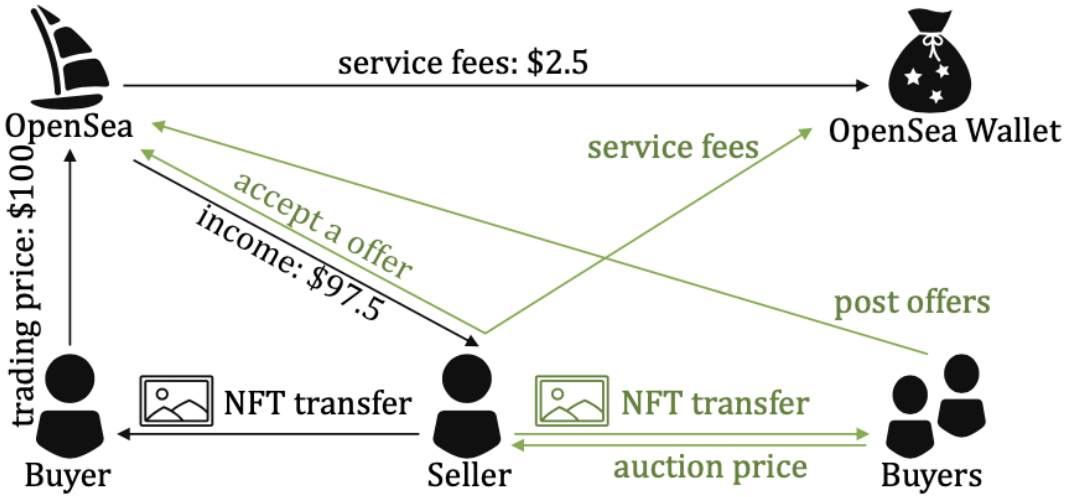
NFT Marketplace

OpenSea

Founded in Dec 2017 by Alex Atallah and Devin Finzer, OpenSea is a generalized marketplace set up to facilitate the trading of NFTs. is by far the largest NFT marketplace by trading volume. It saw about $3 billion in trading volume in September 2021 alone, making up more than 99% of the total market.



OpenSea Website



Two main ways to purchase NFTs on Opensea:
Instant sale (Black) / Auction (Green)

Data Query

BTC

Full-Node (Local)

citp/**BlockSci**

A high-performance tool for blockchain science and exploration

API

Blockchain.com APIs

Build bitcoin apps on top of our APIs for free. **API Terms of Service**

ETH

Full-Node (Local)

ethereum/**web3.py**

A python interface for interacting with the Ethereum blockchain and ecosystem.

API

ETHERSCAN

The Ethereum Block Explorer

Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path

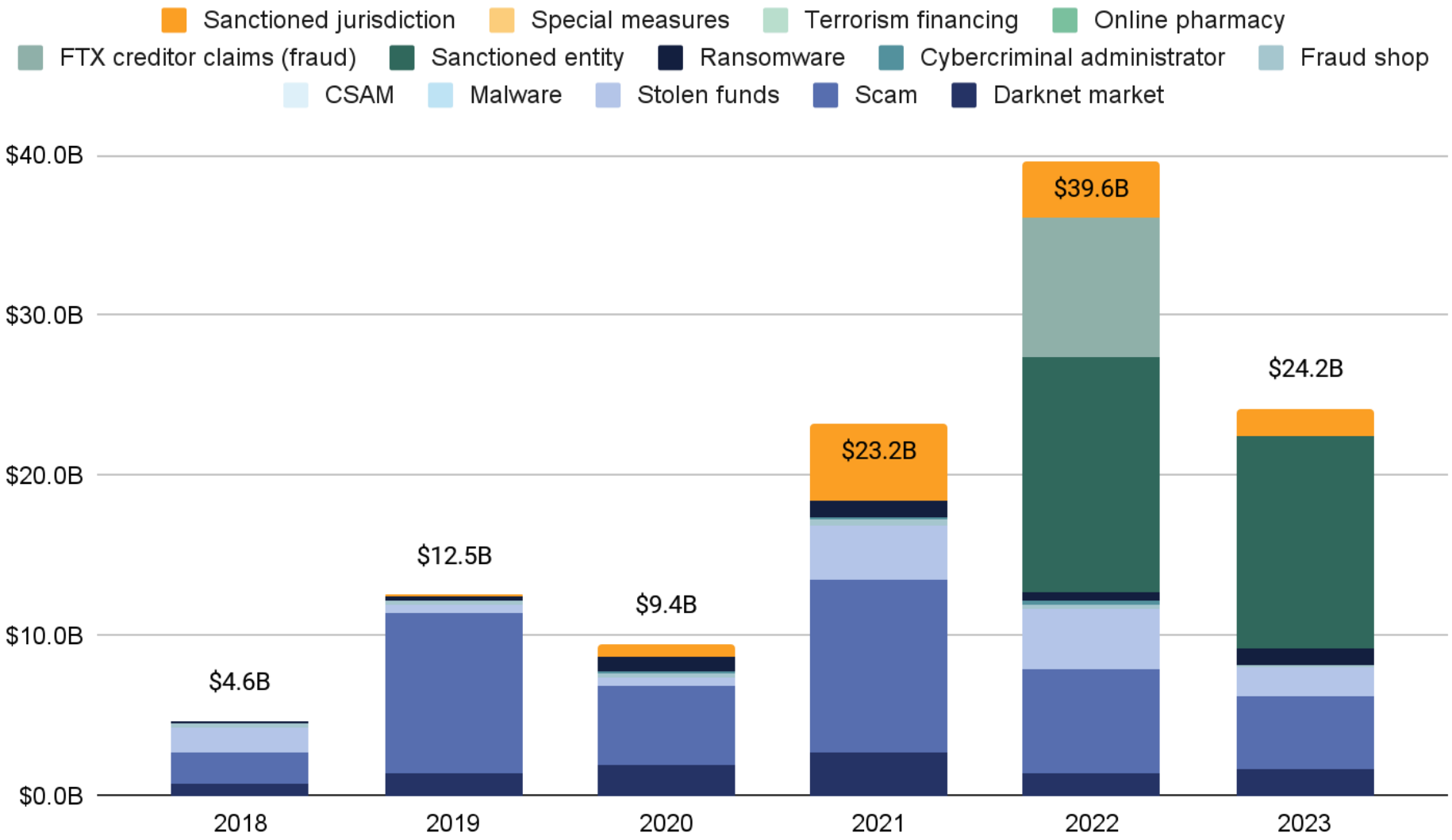
Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path

Crypto-Crime Volume is Tremendous

School of
Computing and
Information Systems

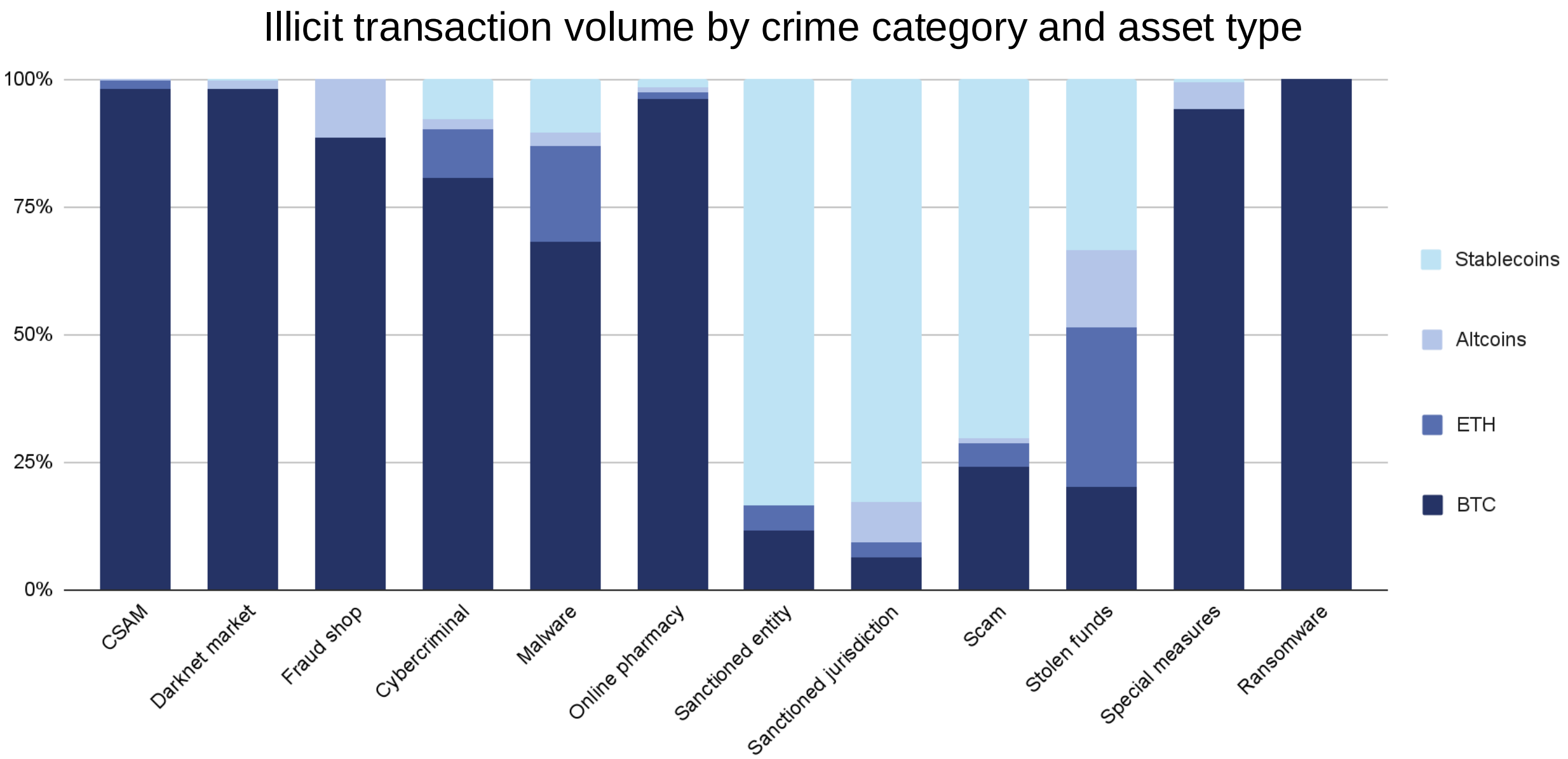
Total cryptocurrency value received by illicit addresses



1. https://go.chainalysis.com/crypto-crime-2024.html

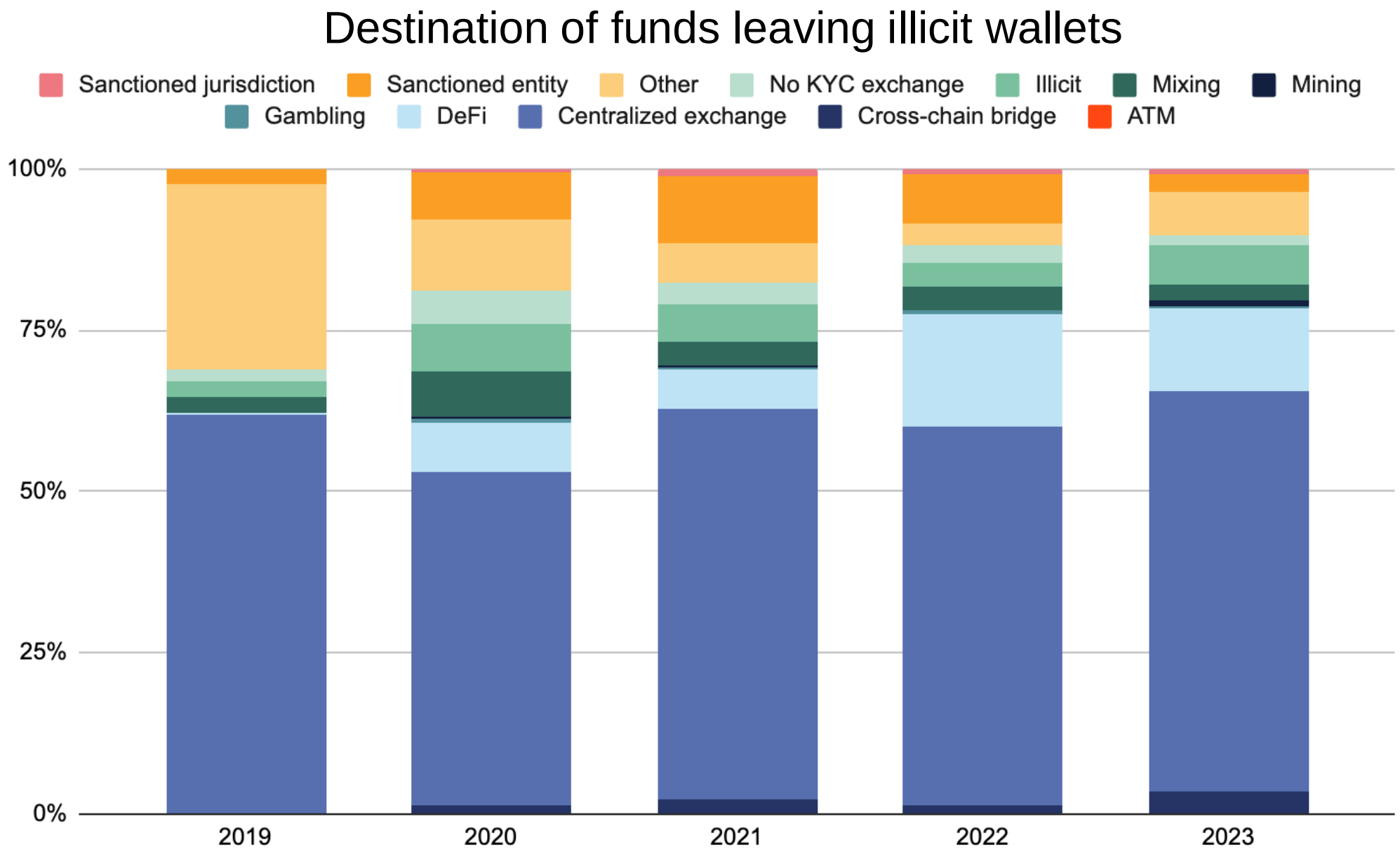
Asset Type of Various Illicit Transactions

School of Computing and Information Systems



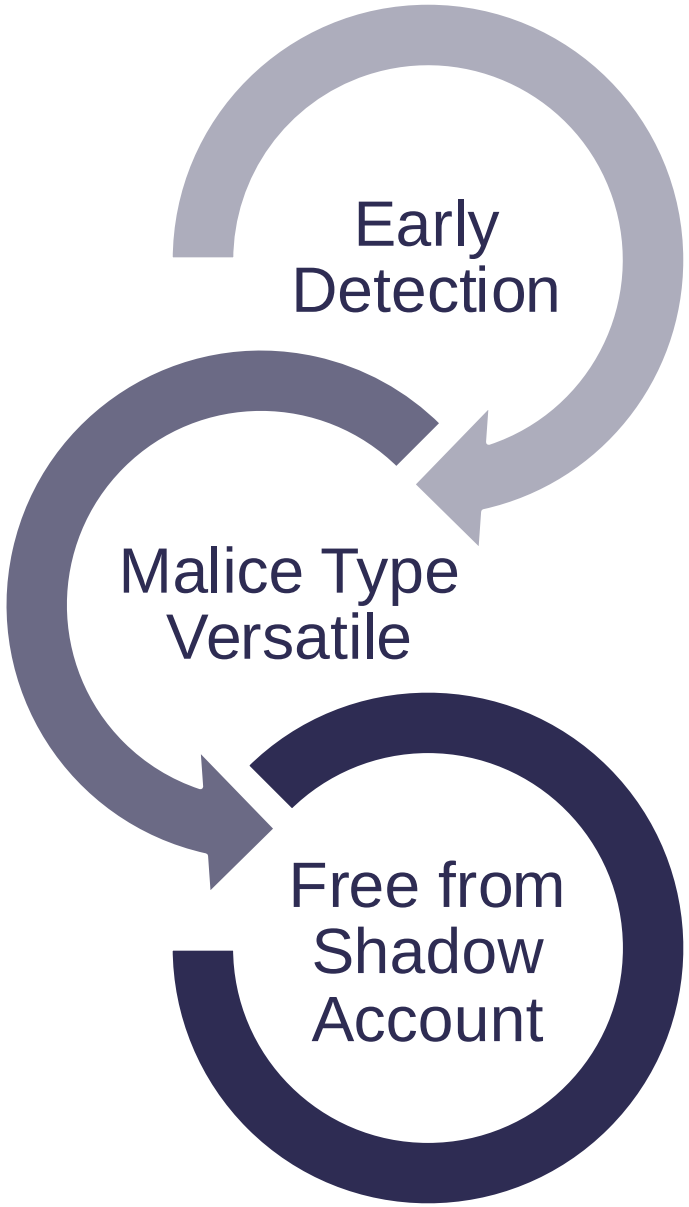
Destination of Illicit Asset

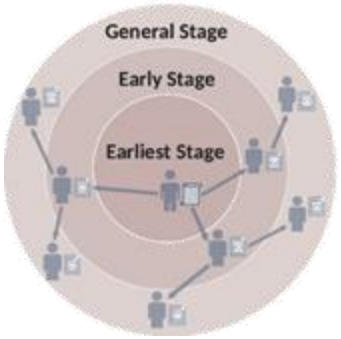
School of
Computing and
Information Systems



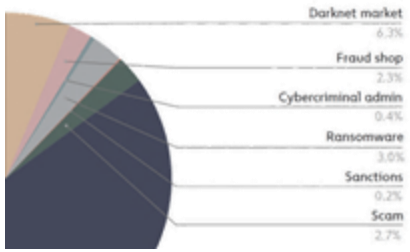
Ideal Model

School of
Computing and
Information Systems

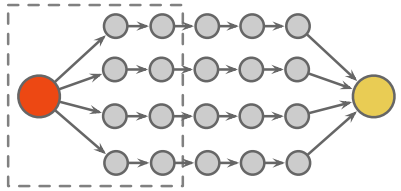




Most malice last for a short duration and cause damage if not be detected in the early stage.



Malice types are constantly evolving. Manually-engineered features for a specific type cannot be generalized to others.

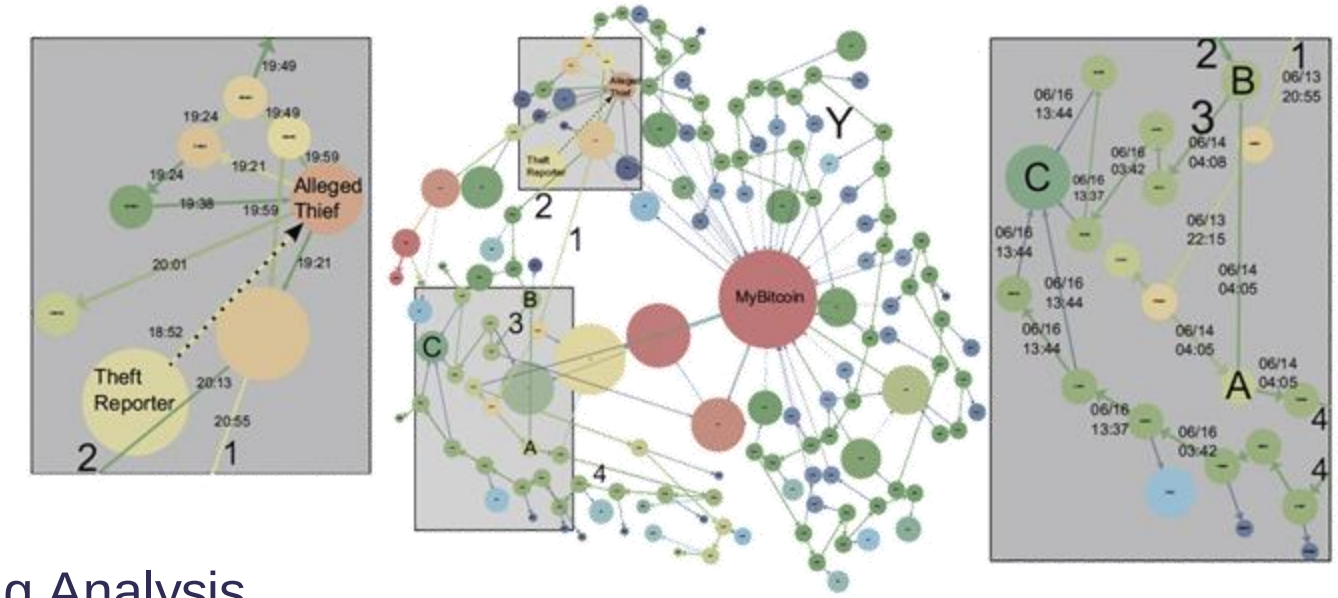


Most address transaction network methods suffer shadow address issues. These addresses introduce extremely long transfer chain.

Current Methods

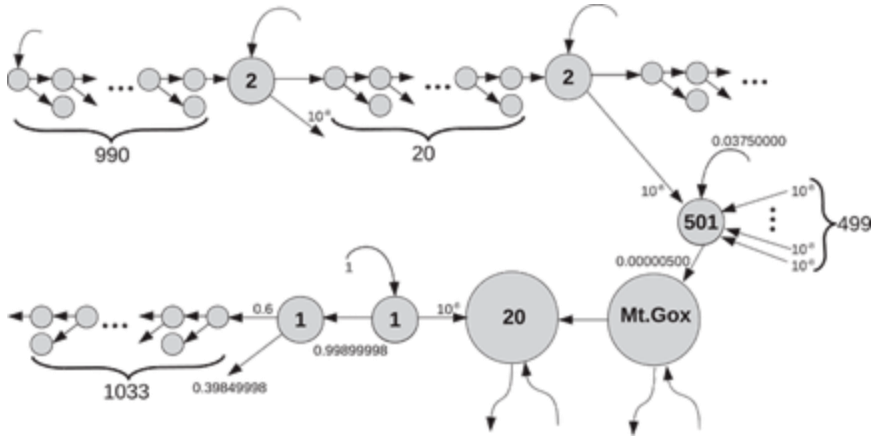
School of Computing and Information Systems

Visualization Based



- Scatter Center
- Hierarchical Transition
- Long Chain
- Self Loop

Tracking Analysis



- Off-Chain Information
- Amount Analysis
- Money Source Tracking
- Temporal Pattern
- Heuristic Rule

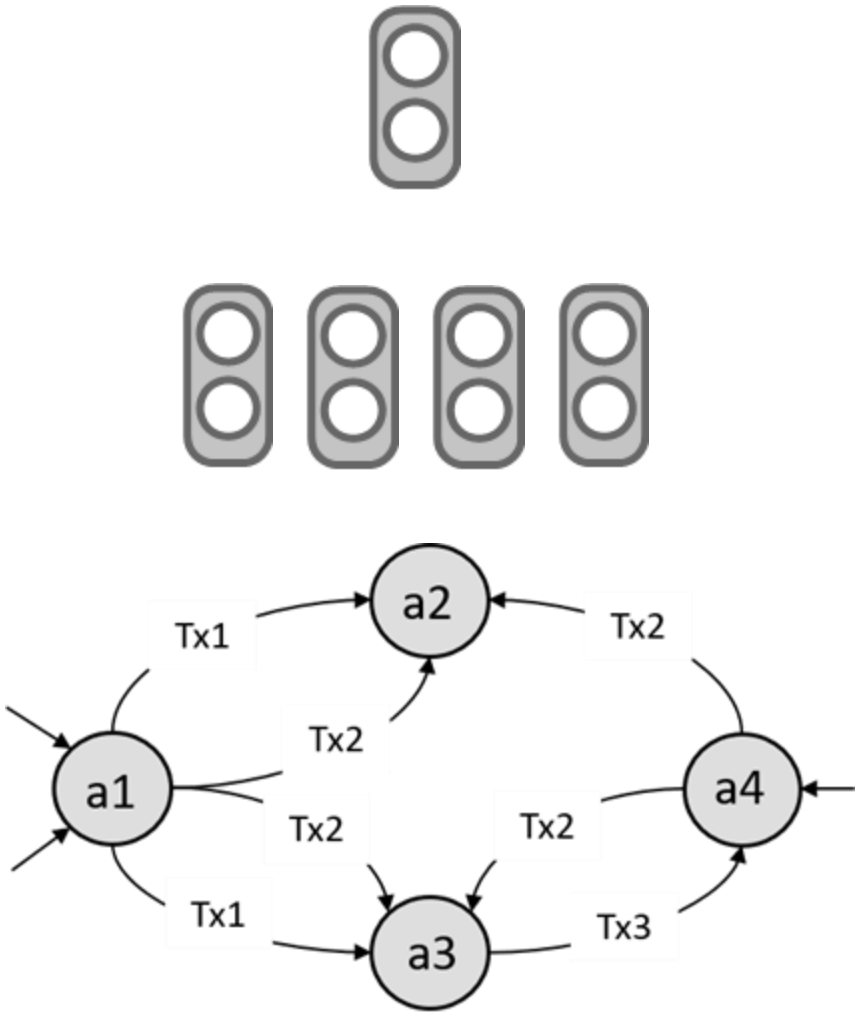
1. Dorit Ron. *"Quantitative Analysis of the Full Bitcoin Transaction Graph"*. FC(2012).

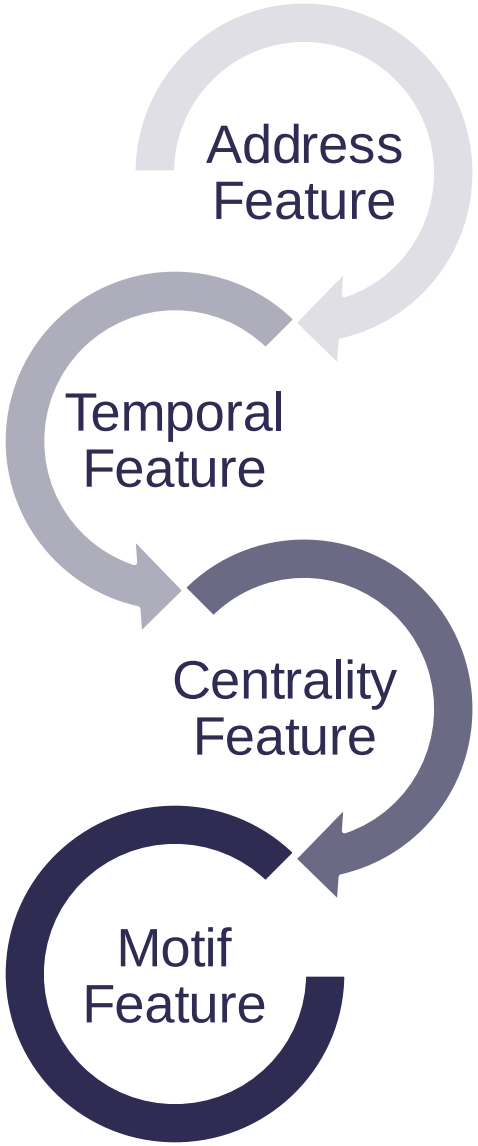
2. Chen, Shijian, et al. "*The dark side of nfts: A large-scale empirical study of wash trading*." Proceedings of the 15th Asia-Pacific Symposium on Internetware. 2024

Current Methods

School of
Computing and
Information Systems

Feature Engineering

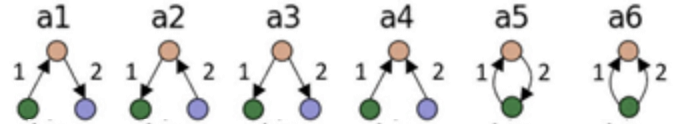


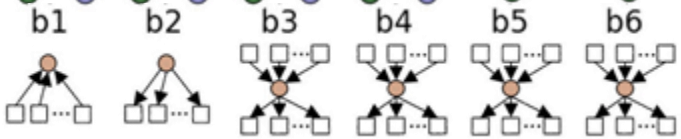


- Total Token Received
- Token Balance
- Number of In/Out Transactions
- Number of Sibling Account

- Number of Active weeks
- Number of entity traded/week
- Number of receive/send days
- Activity period duration

- In/Out Degree
- Betweenness Centrality
- Closeness Centrality
- PageRank

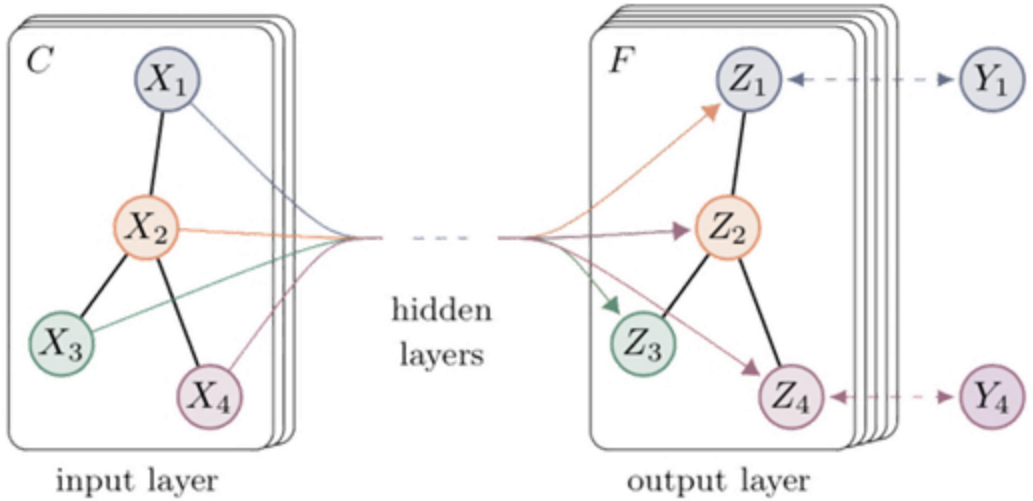
- 

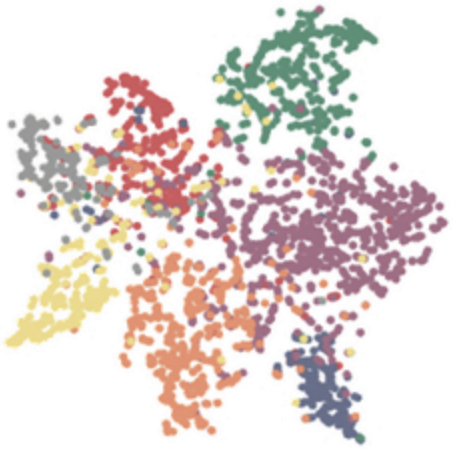
- 

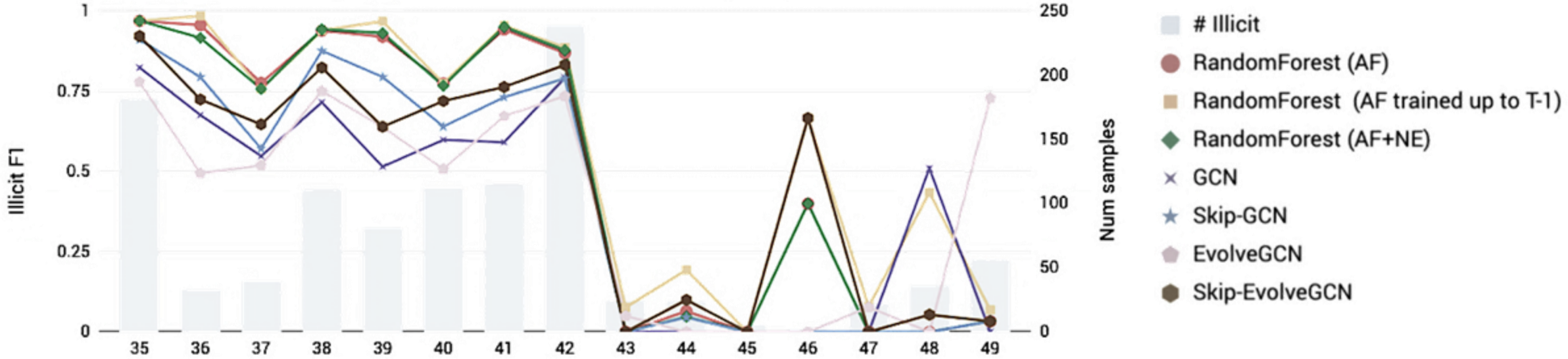
1. Zibin Zheng et.al. *"DenseFlow: Spotting Cryptocurrency Money Laundering in Ethereum Transaction Graphs". The Web Conf* (2024)

Current Methods

Graph Neural Network







1. M. Weber, *et al.* "Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics" SIGKDD Workshop, 2019

2. M. Bartoletti, *et al.* "Data Mining for Detecting Bitcoin Ponzi Schemes" CVCBT, 2018

Current Challenges

Current Challenges

Ineffective For Early Detection

Hack of Binance of May 7, 2019. The path through Chipmixer

All of the transactions from table 1 were made in the time period from 06:41 to 15:17 on 2019-06-13 UTC. Our algorithm allows to determine the relationship between deposit transactions and transactions withdrawing BTC from Chipmixer and belonging to the same entity that made the deposit transaction. Using our algorithm, we found transactions that hackers used to withdraw funds from Chipmixer.

January 22, 2021 02:20 JST

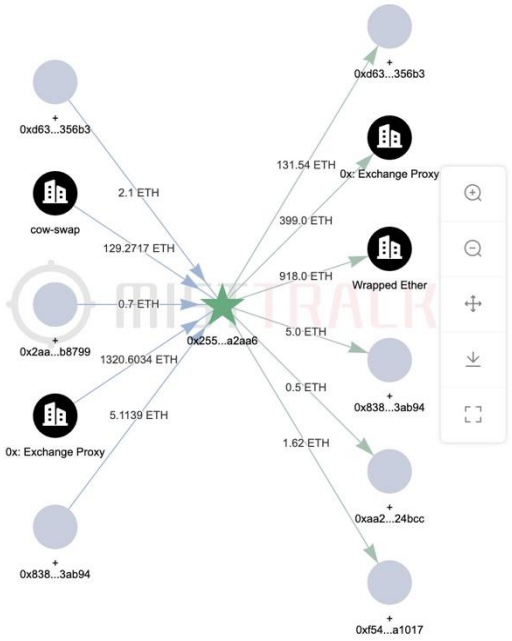
Jan 22, 2021

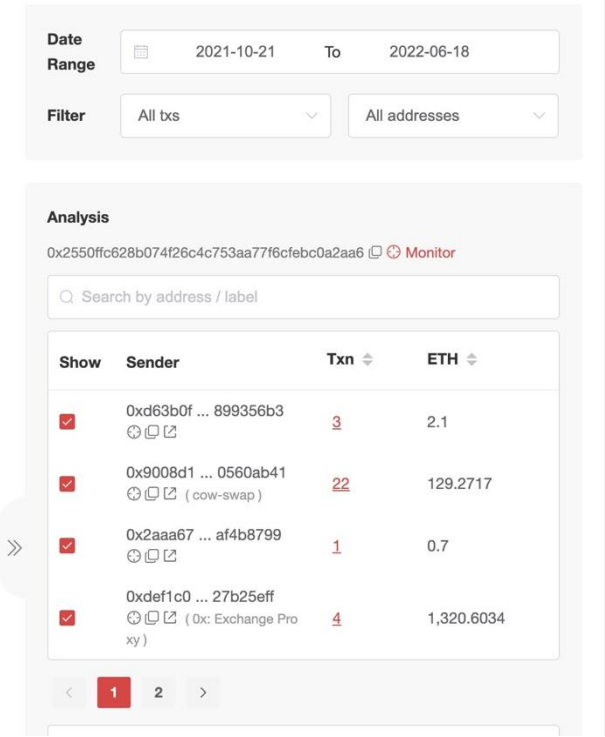
TOKYO -- Police in Japan have identified roughly 30 people for alleged involvement in illegal transactions stemming from 58 billion yen ($530 million at the time) worth of NEM cryptocurrency hacked from the Coincheck exchange three years ago, Nikkei has learned.

The individuals have either been arrested or their cases have been referred to the prosecutors' office, according to a source familiar with the situation.

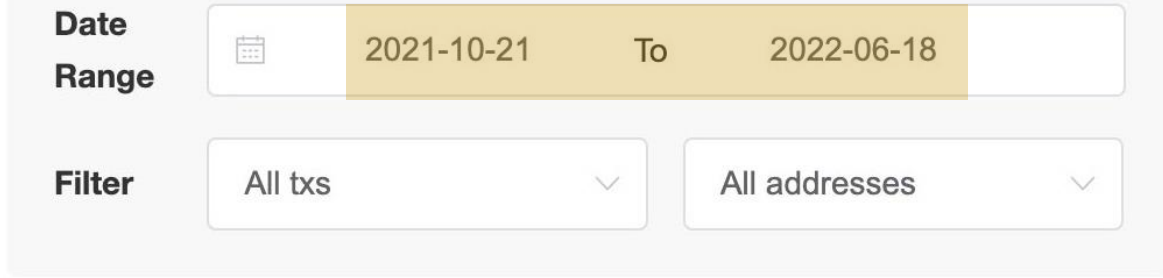
Jan 27, 2018

The 2018 attack on one of Japan's leading cryptocurrency exchanges rattled investors and prompted increased regulatory oversight of the industry.





(Limited Info / Scalability) Issues for GNN



Current Challenges

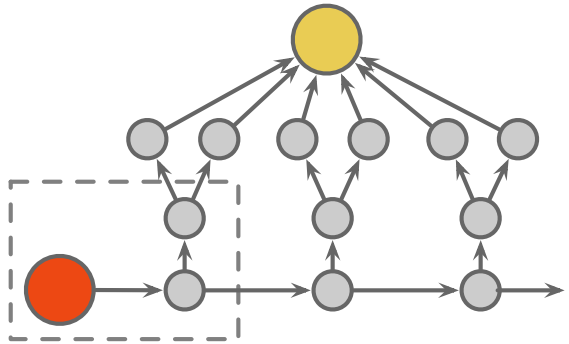
Lack of Versatility

Ref.	Title	Authors	Publication Type	Year
[1]	A novel methodology for HYIP operators' Bitcoin addresses identification	K Toyoda et al.	Journal	2019
[2]	Improving Bitcoin ownership identification using transaction patterns analysis	TH Chang et al.	Journal	2020
[3]	Bitcoin mixing detection using deep autoencoder	L Nan et al.	Conference	2018
[4]	Identifying Bitcoin users using deep neural network	W Shao et al.	Conference	2018
[5]	Detecting mixing services via mining Bitcoin transaction network with hybrid motifs	J Wu et al.	Journal	2021
[6]	Bitcoin theft detection based on supervised machine learning algorithms	B Chen et al.	Journal	2021
[7]	Data mining for detecting Bitcoin Ponzi schemes	M Bartoletti et al.	Conference	2018
[8]	Anomaly detection in Bitcoin information networks with multi-constrained meta path	R Zhang et al.	Journal	2020
[9]	Characterizing and detecting money laundering activities on the Bitcoin network	Y Hu et al.	Journal	2019

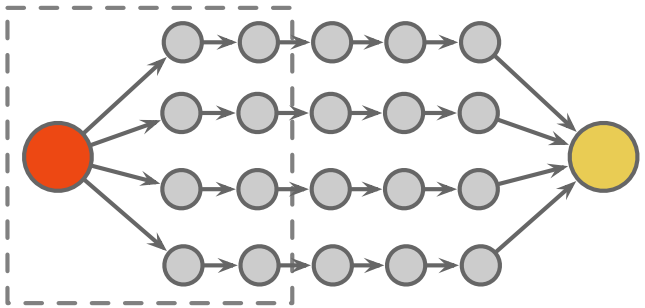
Most are based on interaction analysis with specific entities

Current Challenges

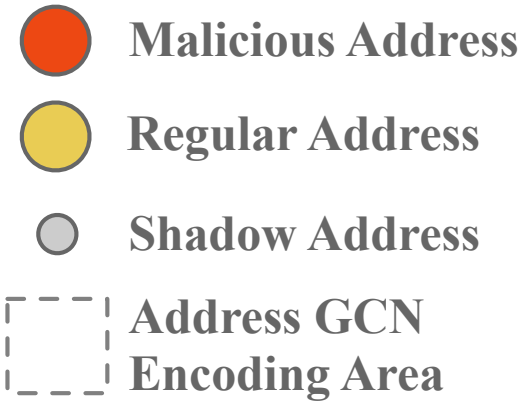
Suffer Shadow Account Issue



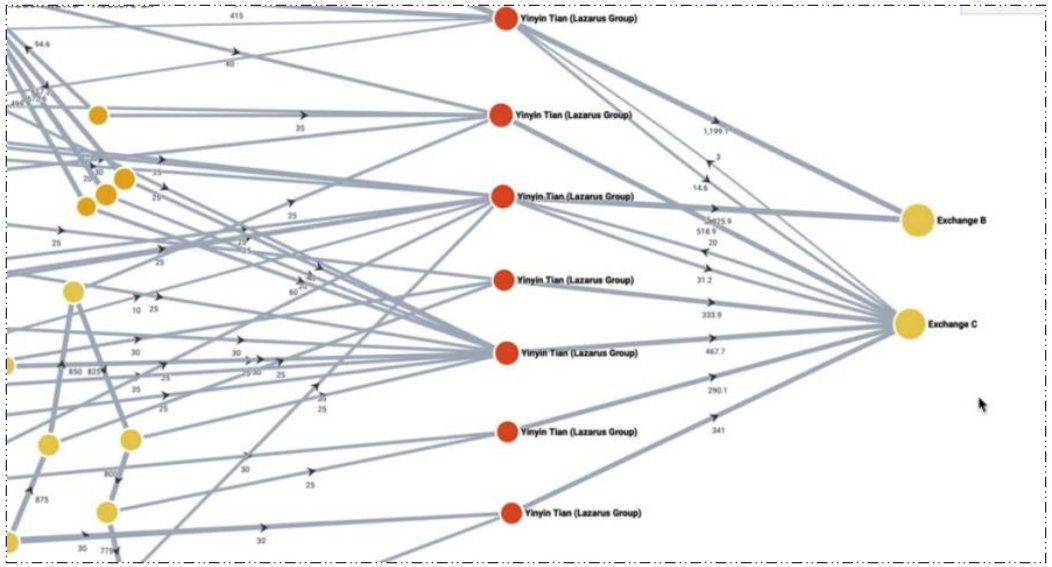
Peeling Chain



Spindle



ELLIPTIC



1. https://www.elliptic.co/blog/elliptic-analysis-bitcoin-bitfinex-theft
2. https://www.elliptic.co/hubfs/Elliptic_Using%20Blockchain%20Analysis%20to%20Mitigate%20Risk%202022%20sanctions.pdf

Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path

Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path

Asset Transfer Path

Asset Transfer Path

School of
Computing and
Information Systems

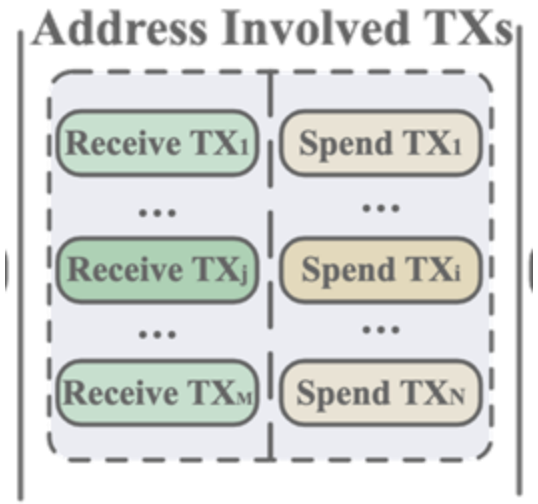
inputs: 3 (0.23154944 BTC)	unique addresses: 3, source transactions: 3		
0. 1AJCCxhKquHMdPW1Lw3RdkqXj9JVbGdQyg	0.16311193 BTC	⇐ 99411a66...	
1. 1CwXs1tbv5Dv3DN1HWLii7thnjJ6DaVXfv	0.01012727 BTC	⇐ e224de23...	
2. 1QGK4uWK43wWkheKyjMJHkUHWnmK8X2J6L	0.05831024 BTC	⇐ 4fff7a1c...	

outputs: 11 (0.23072744 BTC)	unique addresses: 11, spent: 11 in 10 transactions		
0. 37CUkHnZL1fana8DVUBFTb9fj5kmk5amv9	Xapo.com	0.02189381 BTC	b917046e... ⇒
1. 1CDMadTT4mzt2yrEBSKi78XFkdnfaiqy4h	BitoEX.com	0.02212267 BTC	49fdc519... ⇒
2. 3PiAFWjTj8C29PbaKDju6SVb342rZZy1wd	[000004b74b]	0.02232449 BTC	0c13c42f... ⇒
3. 15EqXrJzeFhw63RBACXD5s4DgfcPvCpxvQ	[0004fdeba6]	0.02212267 BTC	2316ed2c... ⇒
4. 12YKAG1ARHFJTBnHu3v3pWgg3dZXMup2i8	(change address)	0.02189381 BTC	47ddc53b... ⇒
5. 12Wsn7Dvi8r97uDRyRaMptkQKCvh8mD591	BitBay.net	0.02211753 BTC	7140669f... ⇒
6. 1P92nSacpTn9VaVTvTkR5akbV5dpeuZU8R	[12ce0a21e7]	0.02197031 BTC	8b51d3f6... ⇒
7. 3JLEyDXPV14oW8N526PLX6tAAcRFde9Cp1	[f041f40b3e]	0.02211496 BTC	a214787d... ⇒
8. 1AQdmPW9GDm5kn576RA4ifYPhFKNjmZnTL	BitBay.net	0.0219782 BTC	185df6c8... ⇒
9. 1NUxerSwgaAZ4LEx2YuT9B7rVFeHcDhDMD	(change address)	0.01001968 BTC	47ddc53b... ⇒
10. 16MhAeYCYu8toj8YqSPe5K1ZqBMp3UjfF9	[974a2667bb]	0.02216931 BTC	0cac457c... ⇒

Asset Transfer Path

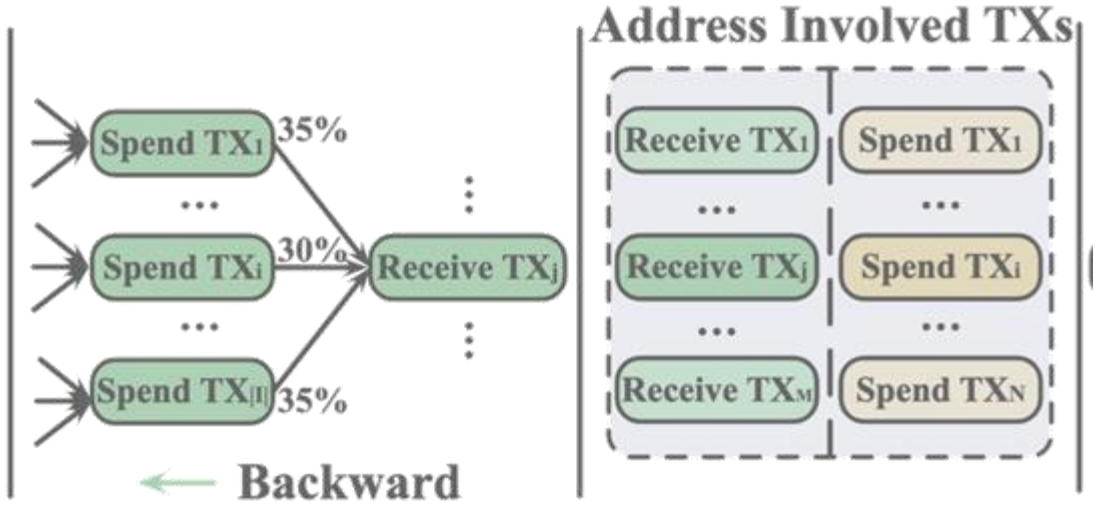
Address 12Wsn7Dvi8r97uDRyRaMptkQKCvh8mD591

date	received/sent	balance	transaction
2019-05-13 04:05:33	+0.00027707	0.00027707	c5fdd21f3e1aa3fe47e6ef30c067258c083995cdc626ffa8de3bc28c9497d40c
2018-04-17 05:53:01	-0.00070748	0.	cd132839ee9171502c861fcca1ab473f72182c116a2eb0ac74c5a80faa128f42
2018-04-16 13:00:59	-0.00009949	0.00070748	d59ec4527a33d182688d5f29b1c751f1f1999229377689264d945d96731e3f65
2017-03-26 18:24:51	+0.00070748	0.00080697	1cb9943096cb1e71f0db154eac4e01f0fa68307c817eb831329cb566bf113f55
2017-02-15 12:25:02	-0.0199187	0.00009949	46b3d8c98a8844c5f9246d48e0010fbf3c58b2b044d22964723b1e20ec0db84b
2017-02-14 18:13:06	+0.0199187	0.02001819	2e7b5370404278348010847 09de62e40c689f9e87b7ea469d6c529316fa2c901
2017-01-25 21:52:11	-0.02211753	0.00009949	7140669f90bb802bdcedd891f8889798c0cb39a5c7f5e99d2e9c44e7c1393b04
2017-01-25 06:08:46	+0.02211753	0.02221702	d8f68e3bf4fea09bd2faa6cee294fecb1502611f1d590cfa4d42a9b2ccb578bd



Asset Transfer Path

inputs: 3 (0.23154944 BTC)	unique addresses: 3, source transactions: 3		outputs: 11 (0.23072744 BTC)	unique addresses: 11, spent: 11 in 10 transactions		
0. 1AJCCxhKquHMdPW1Lw3RdkqXj9JVbGdQyg	0.16311193 BTC	← 99411a66...	0. 37CUkHnZL1fana8DVUBFTb9fj5kmk5amv9	Xapo.com	0.02189381 BTC	b917046e... →
1. 1CwXs1tbv5Dv3DN1HWLii7thnjJ6DaVXfv	0.01012727 BTC	← e224de23...	1. 1CDMadTT4mzt2yrEBSKi78XFkdnfaiqy4h	BitoEX.com	0.02212267 BTC	49fdc519... →
2. 1QGK4uWK43wWkheKyjMJHkUHWnmK8X2J6L	0.05831024 BTC	← 4fff7a1c...	2. 3PiAFWjTj8C29PbaKDju6SVb342rZZy1wd	[000004b74b]	0.02232449 BTC	0c13c42f... →
			3. 15EqXrJzeFhw63RBACXD5s4DgfcPvCpxvQ	[0004fdeba6]	0.02212267 BTC	2316ed2c... →
			4. 12YKAG1ARHFJTBnHu3v3pWgg3dZXMup2i8	(change address)	0.02189381 BTC	47ddc53b... →
			5. 12Wsn7Dvi8r97uDRyRaMptkQKCvh8mD591	BitBay.net	0.02211753 BTC	7140669f... →
			6. 1P92nSacpTn9VaVTvTkR5akbV5dpeuZU8R	[12ce0a21e7]	0.02197031 BTC	8b51d3f6... →
			7. 3JLEyDXPV14oW8N526PLX6tAAcRFde9Cp1	[f041f40b3e]	0.02211496 BTC	a214787d... →
			8. 1AQdmPW9GDm5kn576RA4ifYPhFKNjmZnTL	BitBay.net	0.0219782 BTC	185df6c8... →
			9. 1NUxerSwgaAZ4LEx2YuT9B7rVFeHcDhDMD	(change address)	0.01001968 BTC	47ddc53b... →
			10. 16MhAeYCYu8toj8YqSPe5K1ZqBMp3UjfF9	[974a2667bb]	0.02216931 BTC	0cac457c... →

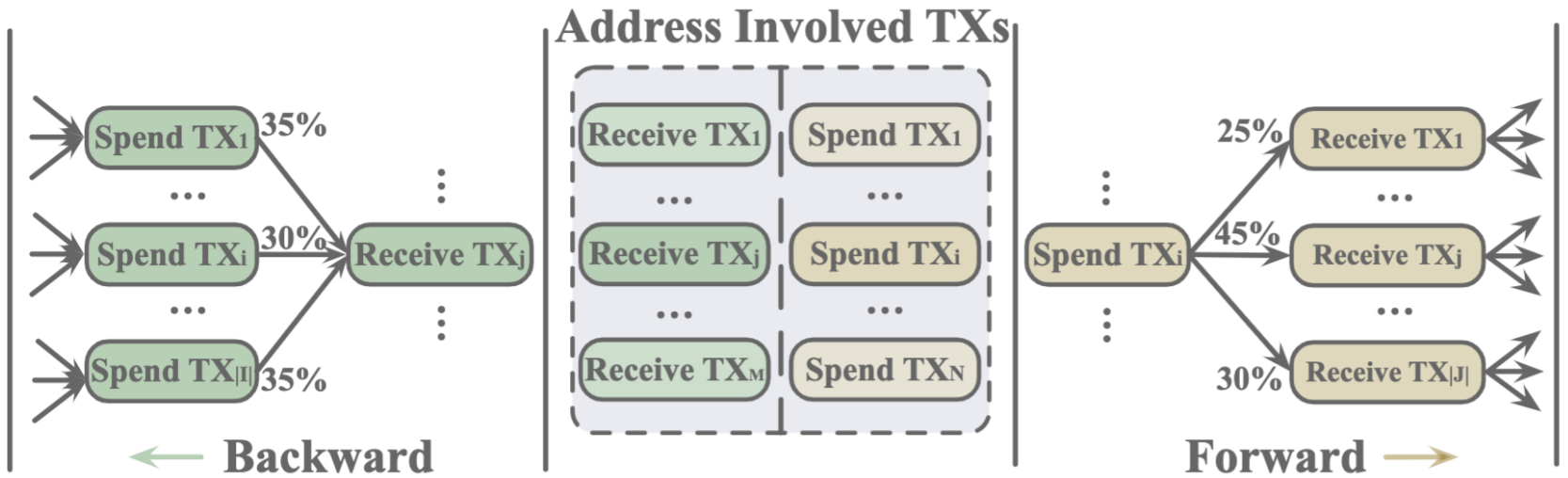


Asset Transfer Path

School of Computing and Information Systems

inputs: 9 (11.93696663 BTC)	unique addresses: 9, source transactions: 9		
0.	1MNT5jyLgGT4LLHLxb4rBR4WwvPRwdm29T	2.67453808 BTC	⇐ e48bfa1c...
1.	16HLm2oN8RQiw59MLf8MVYK5dHLB3Uopif	0.97887336 BTC	⇐ 08ec7df8...
2.	12nqKog2nZYdD2SRpc7cHC1cWHQMHmVbPm	1.08136061 BTC	⇐ 197ce3b0...
3.	1HWC1FL2oiTYSjBx3aFT9wXDjw69bDJYGs	0.48990209 BTC	⇐ e93dc2c1...
4.	1428rdeHpwxubM78iSuQwN3YQEXas2uEX5	0.64988311 BTC	⇐ 54c3203c...
5.	1L3C72iVEGkTe3tSGMo9i8KQp4cLLa12fo	5.00401953 BTC	⇐ 02848604...
6.	14kLma2S3mgkG9gsccujyrwLfPkPFAz1ie	0.54212169 BTC	⇐ 5ebad0a7...
7.	16eDbjoZDfqE4CABfthLtGEzwHUGjzoV82	0.50807124 BTC	⇐ c8d51223...
8.	15fvnjndHFpvw6ZiP2NaNRCFofJPokwBBJ	0.00819692 BTC	⇐ 2fae5f35...

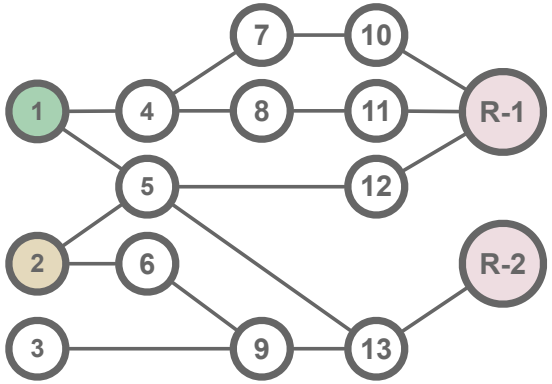
outputs: 3 (11.93656663 BTC)	unique addresses: 3, spent: 3 in 3 transactions			
0.	1DcCYAB5JoCCFittnTehFYhHX5P2uKMpw3	[1e3a5d41d2]	11.88 BTC	3c725cf9... ⇒
1.	15xboUZd3Tmaq3P9fhqxJ2muuZYLw5pWQX	[429c2956d3]	0.04876971 BTC	fc8c6309... ⇒
2.	1G25ESCWCUGxaimhqvd2PNcZ6seSg9Hnzc	[434d20f284]	0.00779692 BTC	49bb8226... ⇒



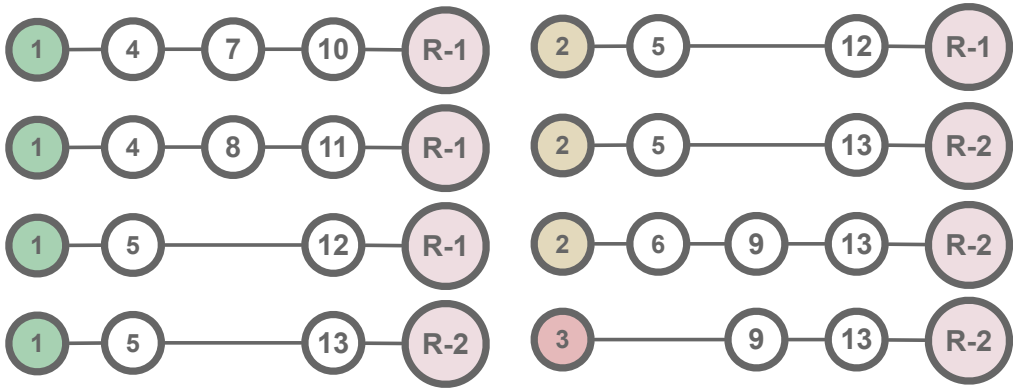
Asset Transfer Path Graph

School of
Computing and
Information Systems

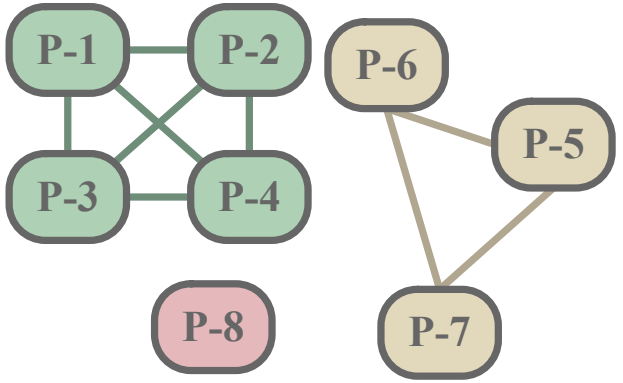
Asset Flow of Receive TX



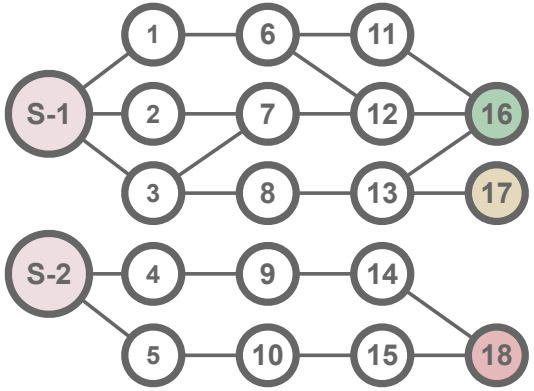
Backward Asset Transfer Path



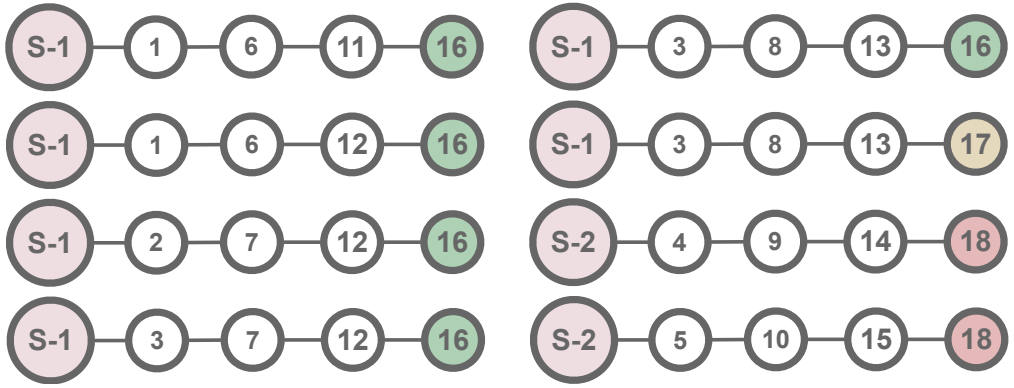
Backward Path Graph



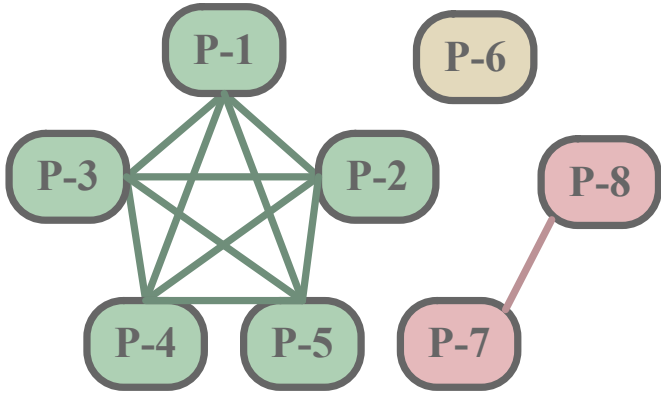
Asset Flow of Spend TX



Forward Asset Transfer Path



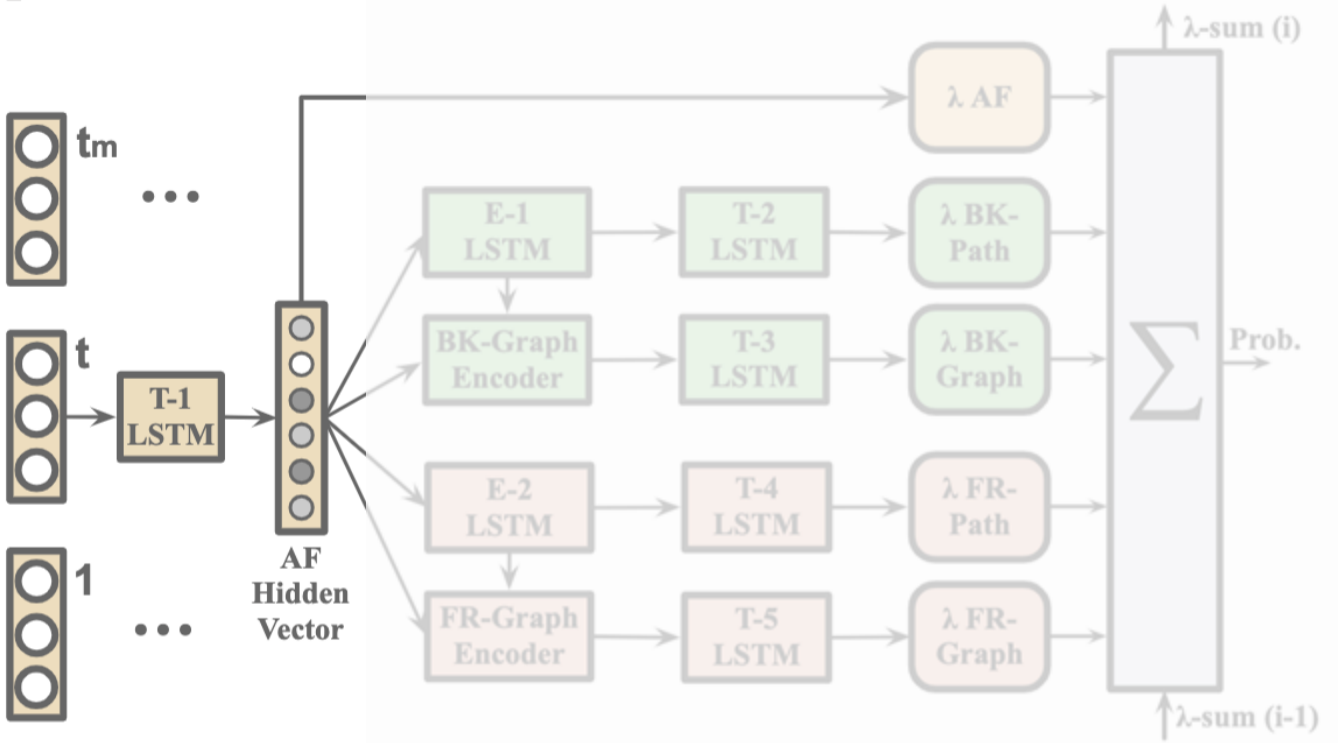
Forward Path Graph



1. Ling CHENG, *et al*. “Evolve Path Tracer: Early Detection of Malicious Addresses in Cryptocurrency” KDD’23

Evolve Path Tracer

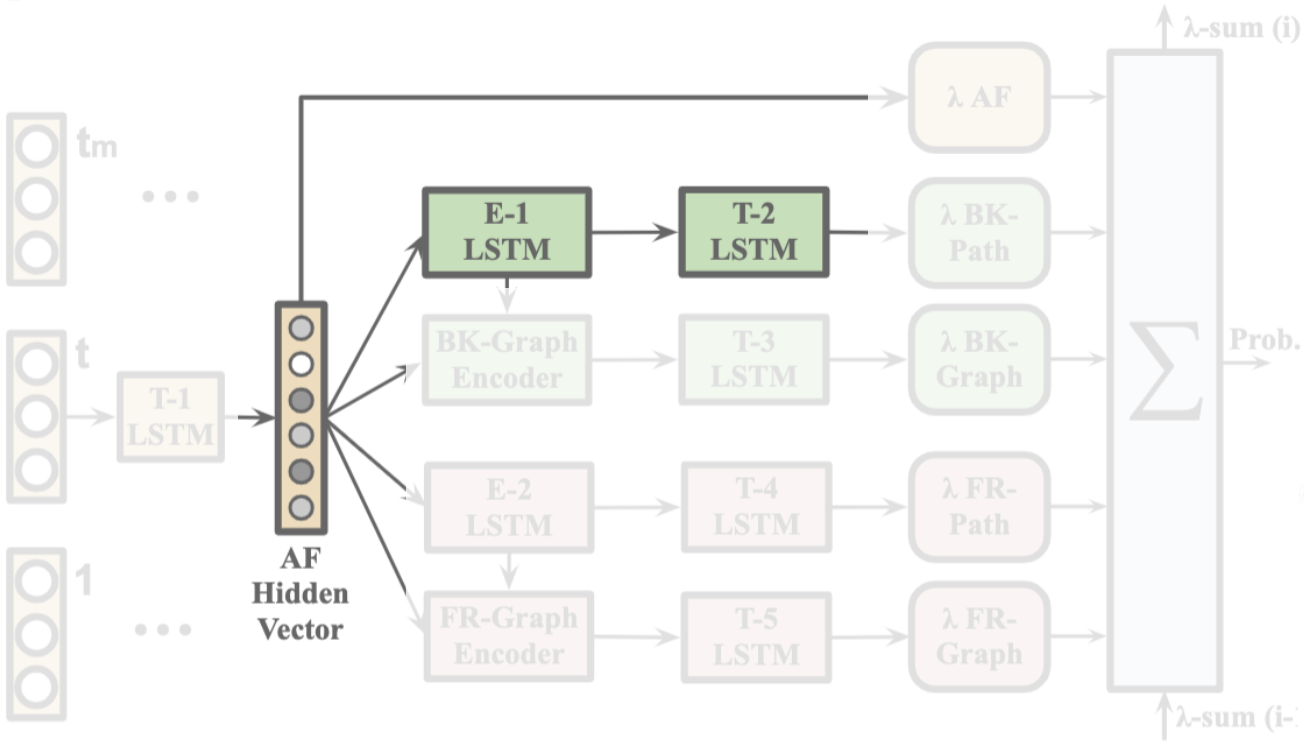
Evolve Path Tracer



- Input current AF into T1 LSTM to get the T1 hidden vector $h_t^{T_1}$.

$$h_t^{T_1}, c_t^{T_1} = \mathrm{LSTM}^{T_1}(f_t^u, h_{t-1}^{T_1}, c_{t-1}^{T_1})$$

Evolve Path Tracer



- Concatenate $h_t^{T_1}$ with $h_{t-1}^{T_2}$ to generate parameters for E1-LSTM, encodes paths to path vectors.

$$h_j^{E_1}, c_j^{E_1} = \mathrm{LSTM}^{E_1}([h_t^{T_1} || h_{t-1}^{T_2}], h_{j-1}^{E_1}, c_{j-1}^{E_1})$$

Evolve Kernel

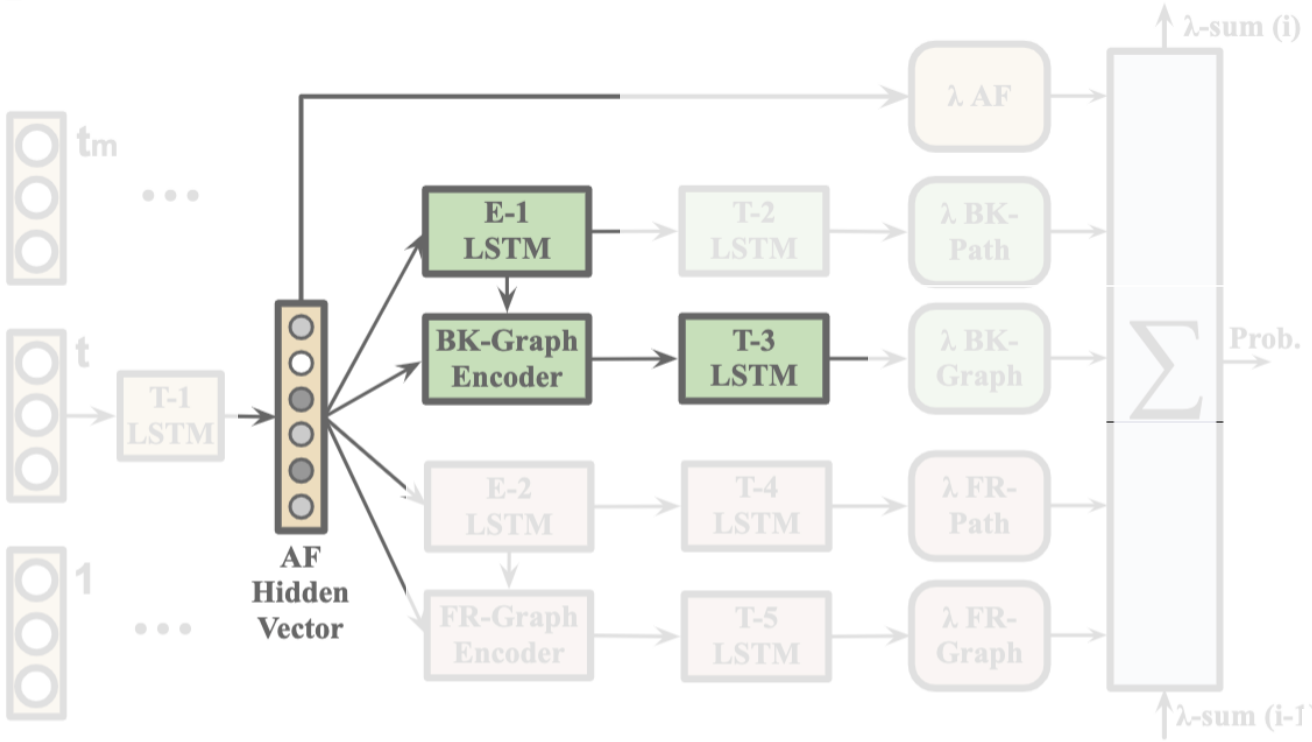
- Then weighted sum them to update T2-LSTM $h_t^{T_2}$

$$a_{i,t}^j = W^{a,j} \tanh(W^{p,u}[f_{i,t}^p || h_t^{T_1}]); \quad \alpha_{i,t}^j = \mathrm{Softmax}(a_{i,t}^j),$$

$$\hat{f}_t^p = ||_{j=1}^{M^p} \hat{f}^{p,j}; \quad \hat{f}_t^{p,j} = \sum_{i=1}^{N_{E_1}} \alpha_{i,t}^j f_{i,t}^p,$$

$$h_t^{T_2}, c_t^{T_2} = \mathrm{LSTM}^{T_2}(\hat{f}_t^p, h_{t-1}^{T_2}, c_{t-1}^{T_2})$$

Evolve Path Tracer



- Concatenate $h_t^{T_1}$ with $h_{t-1}^{T_3}$ to generate parameters for BK-Graph Encoder to update path vectors with graph info. Similarly, weighted sum them to update T3-LSTM $h_t^{T_3}$.

Evolve Kernel

$$H_t^g = [h_t^{T_1} || h_{t-1}^{T_3}],$$

$$f_t^g = \sigma(\tilde{\mathcal{D}}^{-\frac{1}{2}} \tilde{\mathcal{A}} \tilde{\mathcal{D}}^{-\frac{1}{2}} (f_t^p W^g H_t^g)),$$

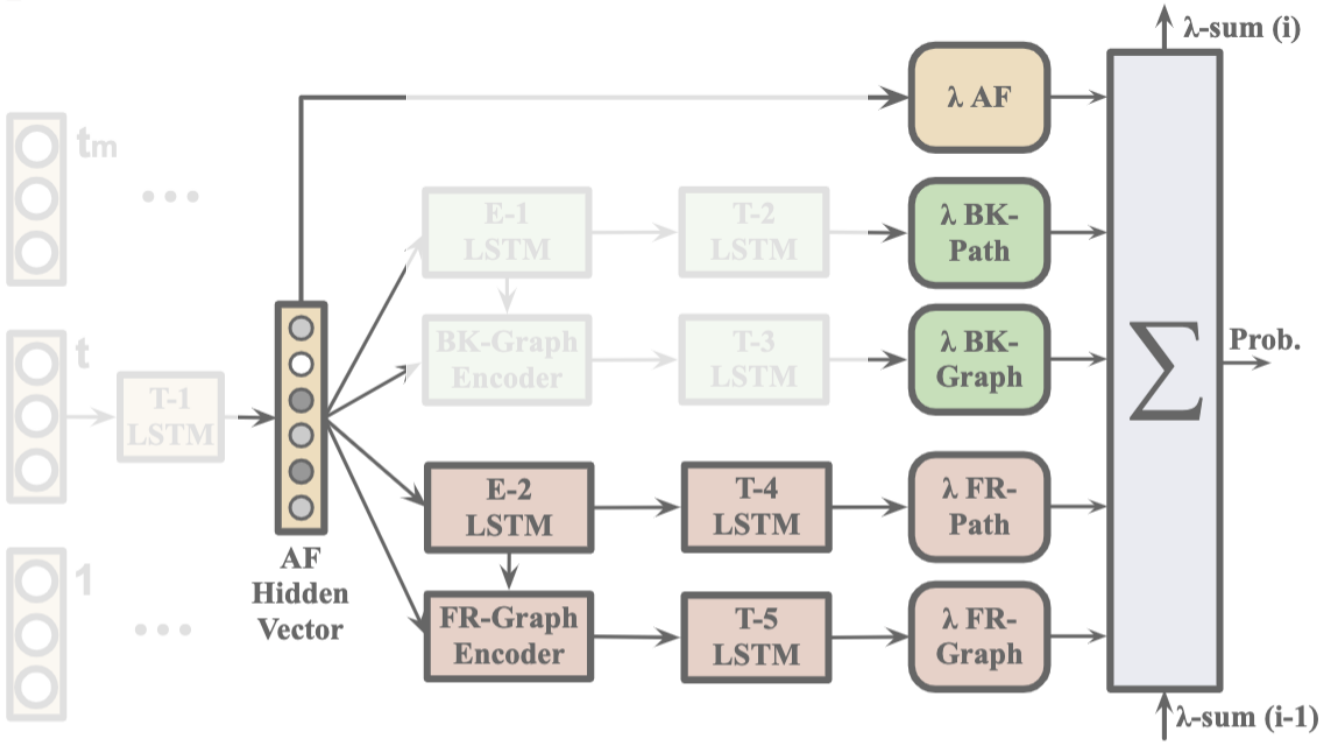
$$\tilde{\mathcal{A}} = \mathcal{A} + \mathcal{I}; \quad \mathcal{A}_{i,:,j} = (W^e H_t^g) S_{i,j},$$

$$\tilde{\mathcal{D}} = \mathrm{diag}(\sum_j (A_{i,j} + \mathcal{I}_{i,j})),$$

$$h_t^{T_3}, c_t^{T_3} = \mathrm{LSTM}^{T_3}(\hat{f}_t^g, h_{t-1}^{T_3}, c_{t-1}^{T_3}).$$

Evolve Path Tracer

- Same processing for the forward branch, use $h_t^{T_1} \sim h_t^{T_5}$ to predict the current label for the given address.

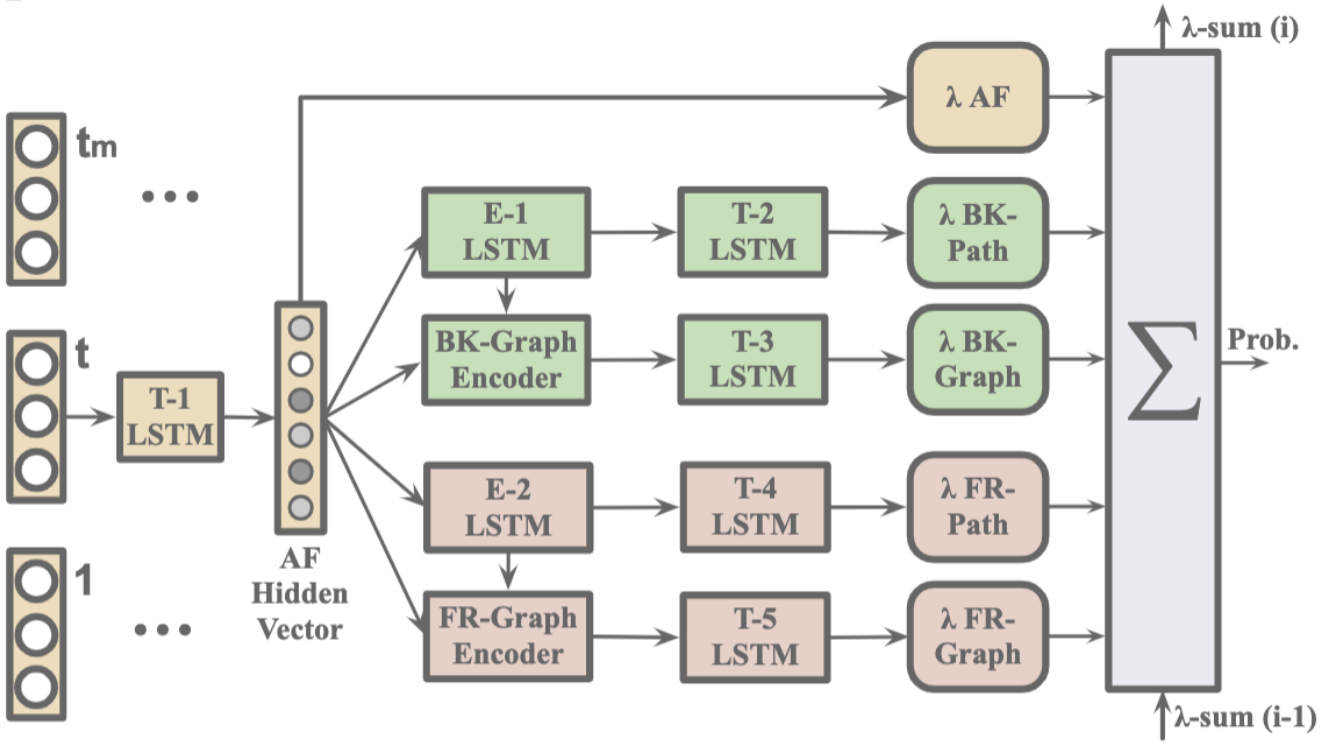


$$\lambda_{j,t} = \tanh(W_{T_j}^{hz} h_t^{T_j}),$$

$$\hat{y}^t = \exp(-\mathrm{ReLU}(\sum_{i=1}^{t}\sum_{j=1}^{5} \lambda_{j,i})),$$

Evolve Path Tracer

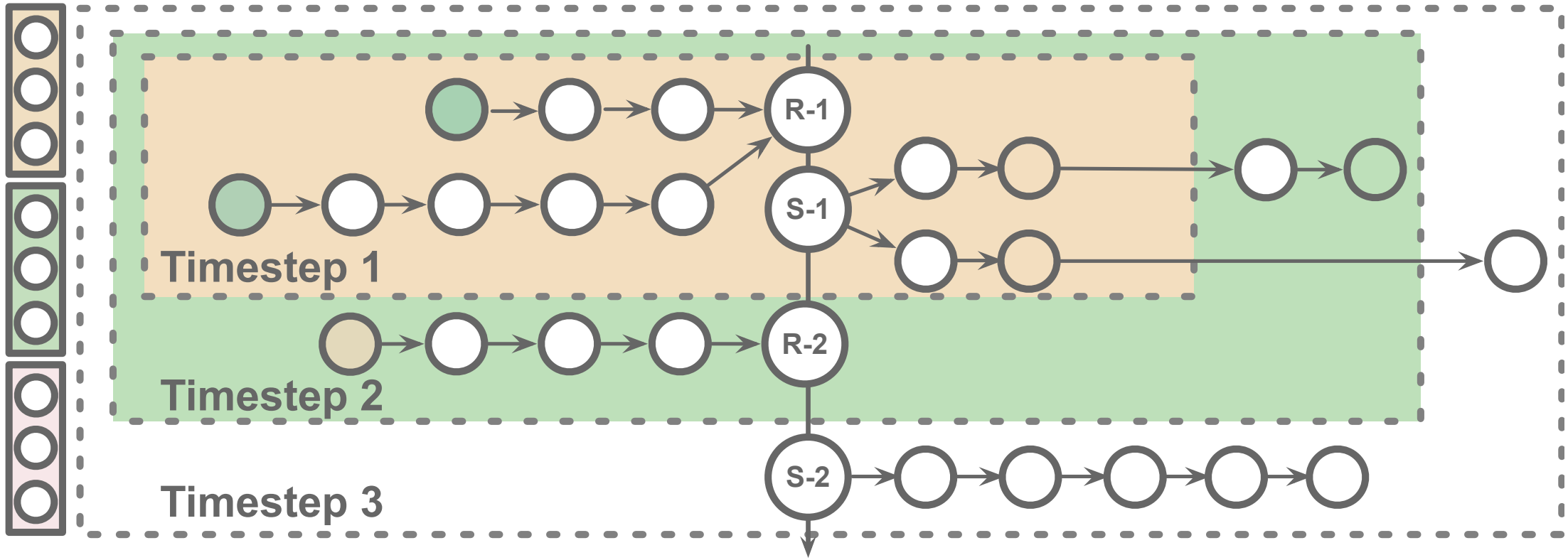
School of
Computing and
Information Systems



- Input current AF into T1 LSTM to get the T1 hidden vector h_t^{T1}.
- Concatenate h_t^{T1} with previous hidden vectors of T2 and T3 LSTM (h_{t-1}^{T2}, h_{t-1}^{T3}) to generate parameters for E1-LSTM and BK-Graph Encoder.
- E1-LSTM encodes paths to path vectors. Then weighted sum them to update T2-LSTM h_t^{T2}.
- BK-Graph Encoder updates path vectors with graph info. Similarly, weighted sum them to update T3-LSTM h_t^{T3}.
- Same processing for the forward branch, use $h_t^{T1} \sim h_t^{T5}$ to predict the current label for the given address.

Asset Transfer Path

Evolution of Asset Transfer Path



Experiments

- Basic Settings

Table 1: Dataset Statistics

Type	Definition	Posi.	Nega.	P/N(%)
H	Hack and steal tokens	302	6582	4.03
R	Encrypt data for ransoms	3224	21100	15.28
D	Illegal BTC darknets	5838	109937	5.31

$$F1^E = \frac{\sum_{i=1}^{N} F1_i/\sqrt{i}}{\sum_{i=1}^{N} 1/\sqrt{i}},$$

$$F1^C = \frac{\sum_{i=1}^{N-1} \sqrt{i} \times F1_i \times \mathbb{1}_{y_c}(y_i)}{\sum_{i=1}^{N-1} \sqrt{i}},$$

- Ablation Study

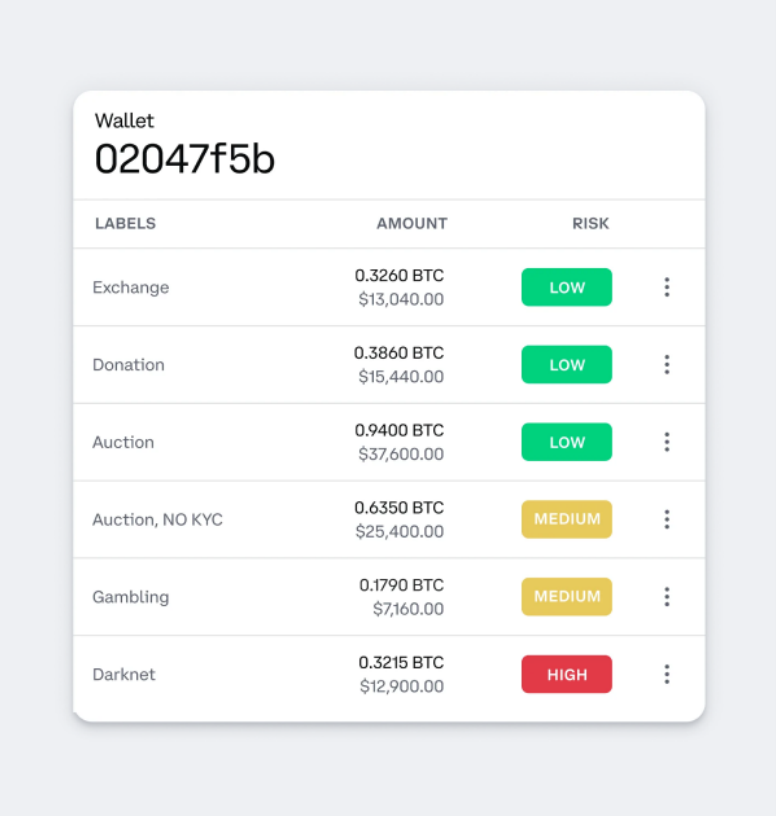
	Model	*Acc.*	*Prec.*	*Rec.*	$F1^E$	$F1^C$
H	AF	0.920	0.309	**0.590**	0.389	0.412
	+Path	0.954	0.537	0.546	0.509	0.538
	+Graph	**0.965**	**0.686**	0.476	0.545	0.559
	+Evolve	0.961	0.606	0.553	**0.551**	**0.576**
R	AF	0.911	0.710	0.632	0.626	0.628
	+Path	0.929	0.727	0.805	0.760	0.765
	+Graph	0.927	0.696	**0.875**	0.773	0.776
	+Evolve	**0.937**	**0.735**	0.871	**0.795**	**0.798**
D	AF	0.961	0.619	0.571	0.611	0.604
	+Path	0.961	0.611	0.693	0.649	0.650
	+Graph	0.960	0.586	**0.804**	0.678	0.678
	+Evolve	**0.963**	**0.626**	0.758	**0.685**	**0.685**

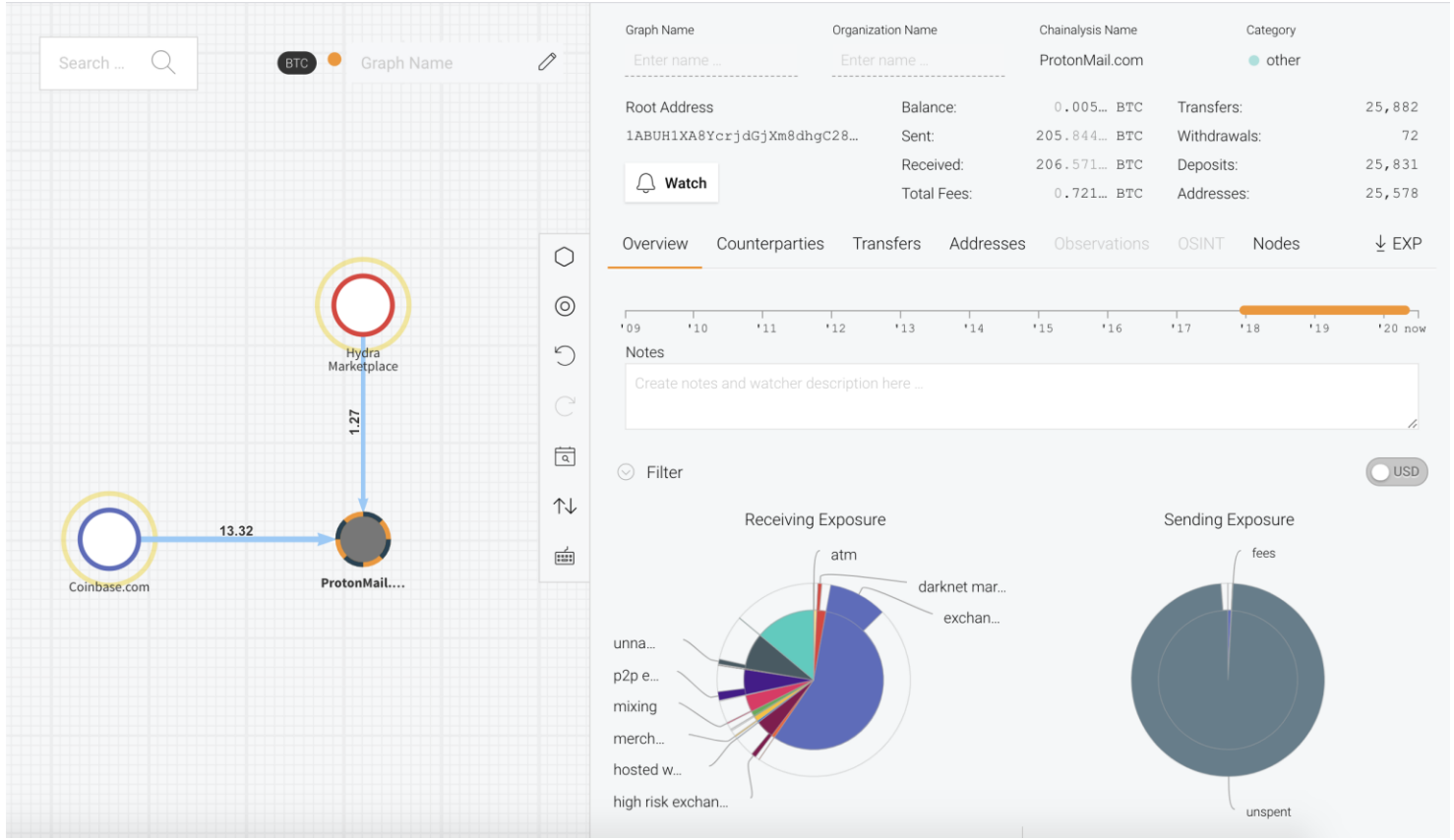
Analysis of Address Intention (Illicit Activity)

Current Challenges

School of Computing and Information Systems

Lack Of Interpretability

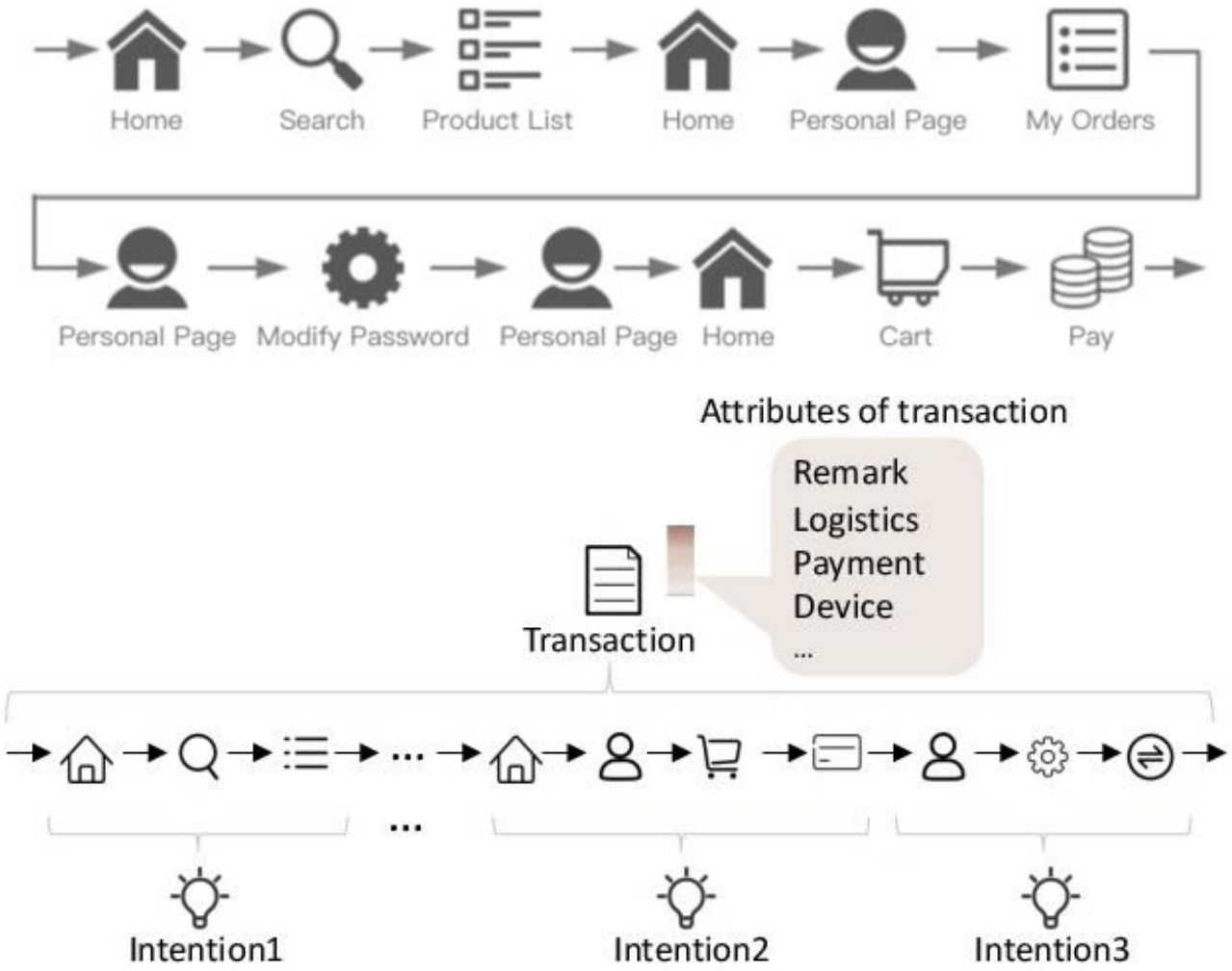


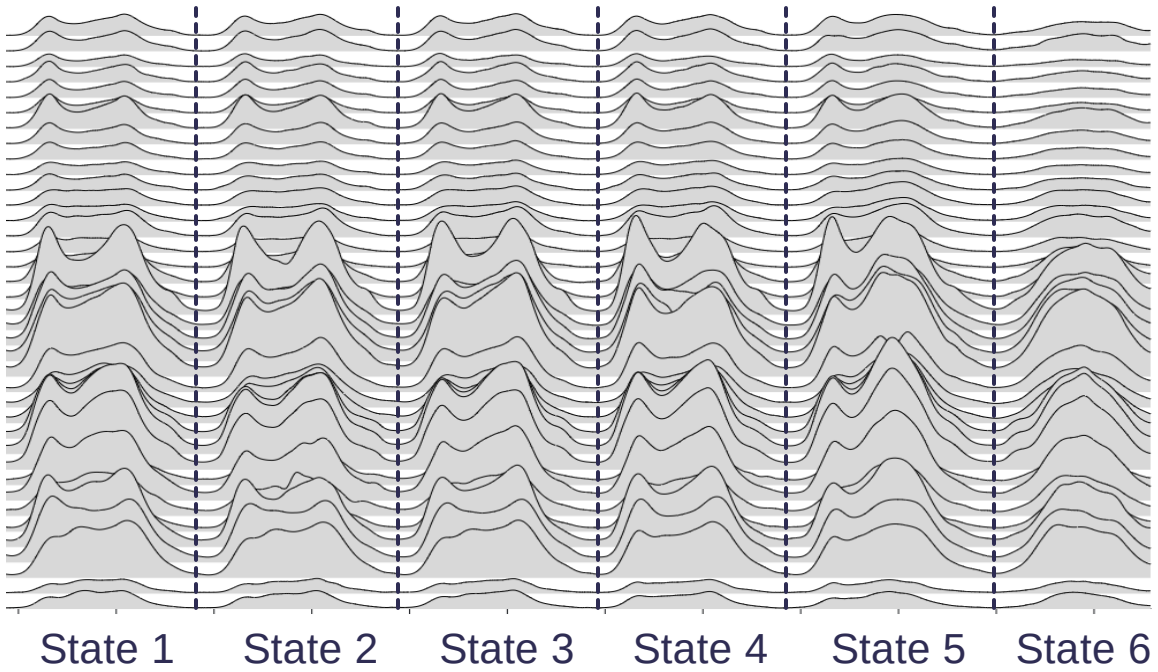


Investors need to tell real creditable projects from frauds.
Current models can hardly offer insights for their predictions.

Intention Monitor

School of
Computing and
Information Systems



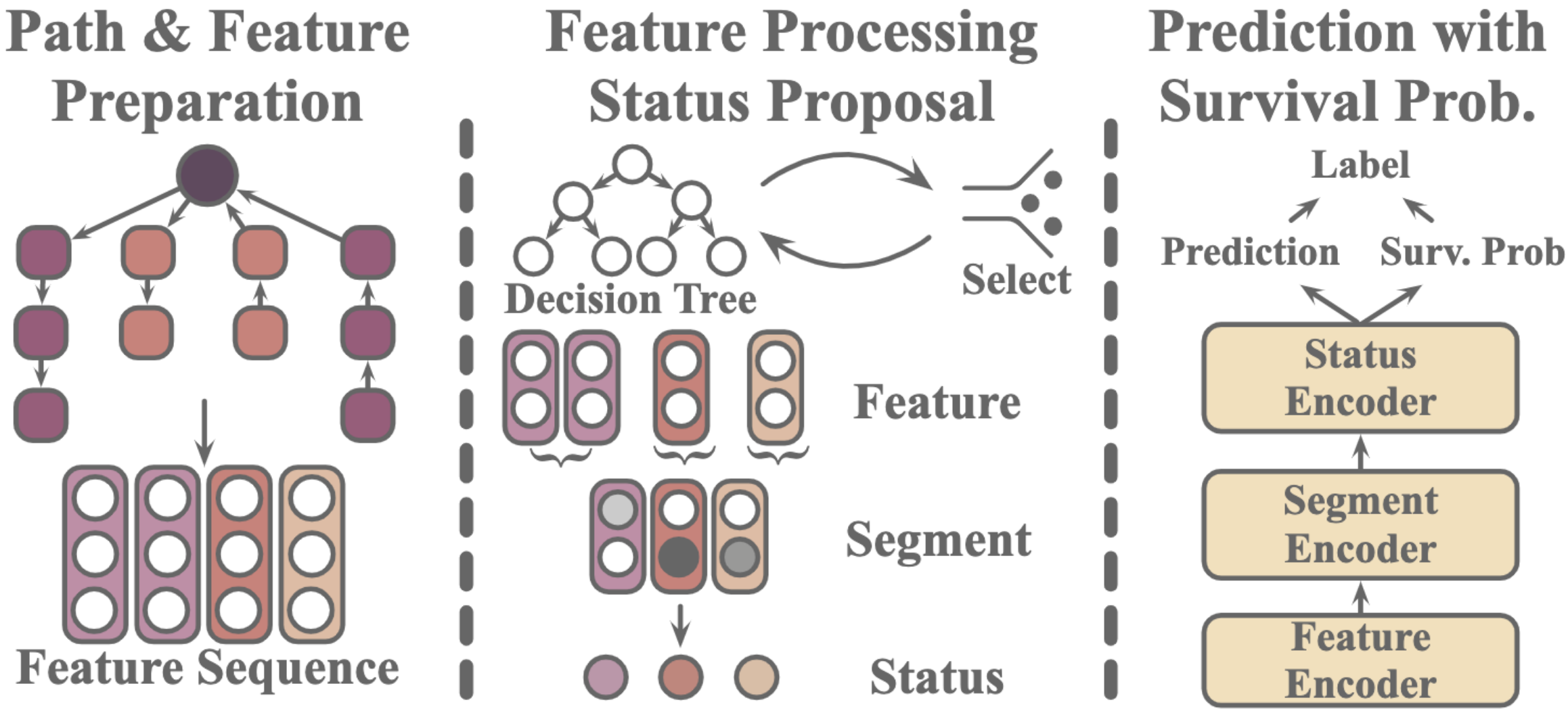


Status → Action → Intention

1. Liu, Can, et al. "Fraud transactions detection via behavior tree with local intention calibration." SIGKDD. 2020

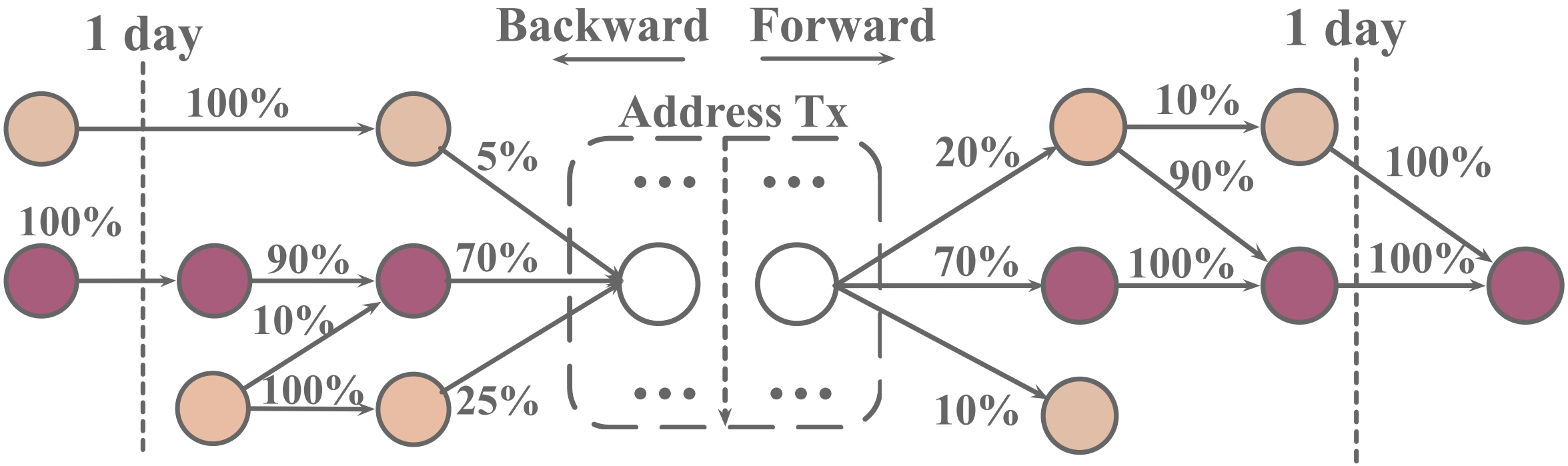
2. Liu, Can, et al. "Intention-aware heterogeneous graph attention networks for fraud transactions detection." SIGKDD. 2021

Outline



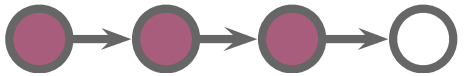
1. Ling CHENG, *et al*. "From Asset Flow to Status, Action, and Intention Discovery: Early Malice Detection in Cryptocurrency" TKDD 2023

Asset Transfer Path

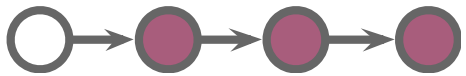


Asset Transfer Path

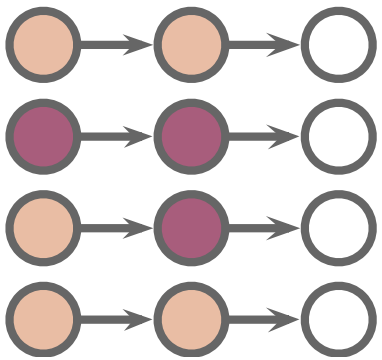
LT-BK



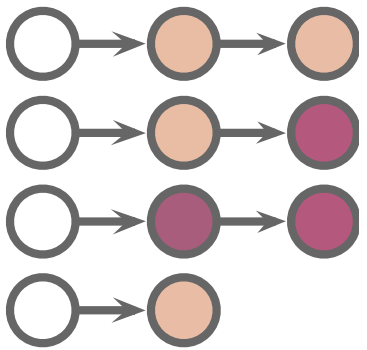
LT-FR



ST-BK

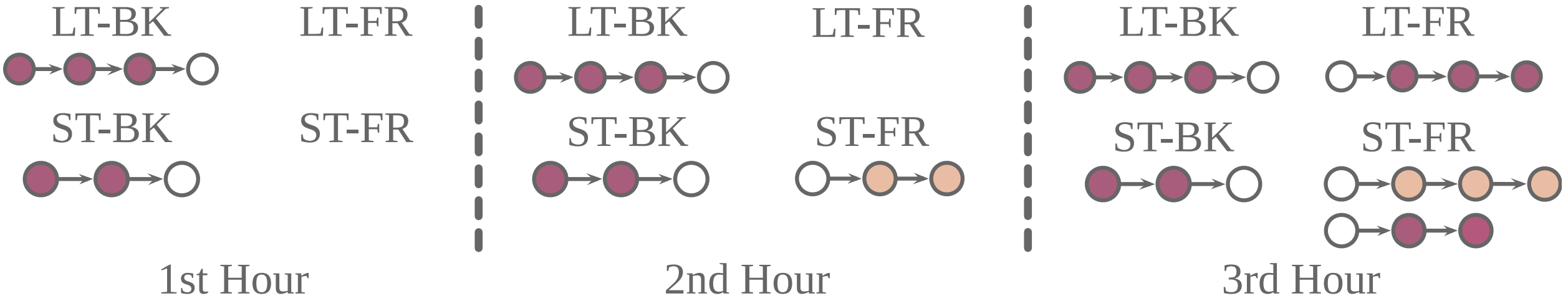


ST-FR



Asset Transfer Path

Evolution of Asset Transfer Path

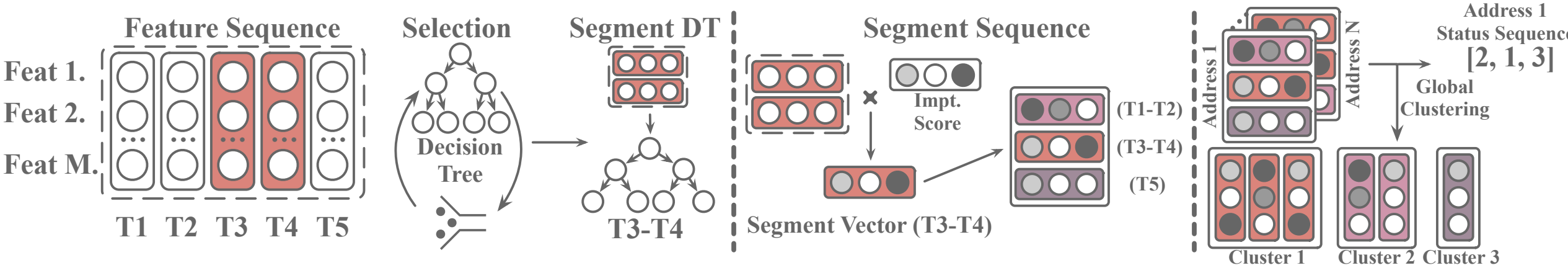


Asset & Path Features

Feature Type	Aspect	Feature	Num.	Complement
Address	Balance	Balance	1	None
	Tx Count	Number of spend (receive) tx (by now/recent one hour)	4	None
		Ratio of spend tx to recieve tx (by now/recent one hour)	2	None
	Tx Frequency	Max spend (receive) tx number per hour	2	None
	Abnormal Tx	Number of spend (receive) tx with 0 amount	2	None
	Temporal Info	Time of max hourly spend (receive) tx number	2	None
		Time difference between max hourly spend and receive	1	None
	Activity	Active hour number and active rate	2	None
Path (LT/ST)-(BK/FR)	Path-Count	Path number	1	None
	Path-Length	Hop (height)-length	2	Min,Max,Std
	Tx Amount	Max (min) input (output) amount	4	Min,Max,Std
	Tx Structure	Max (min) input (output) tx number	4	Min,Max,Std
	Connectivity	Path's max (min) activation score	2	Min,Max,Std

Intention Monitor

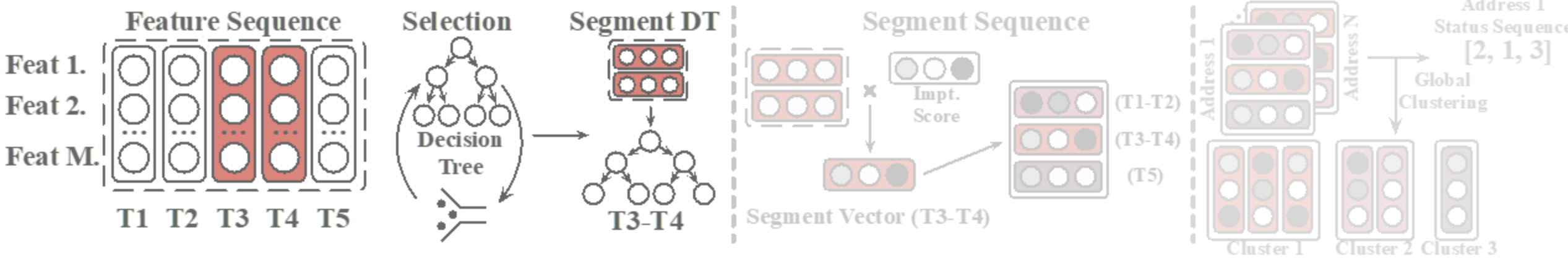
Overview of Intention Monitor



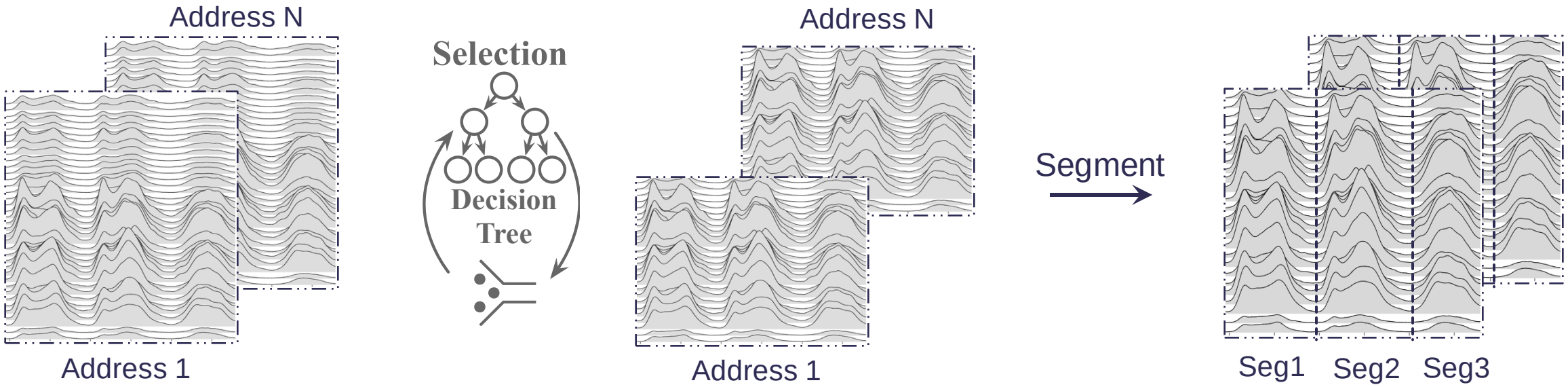
How to propose status from temporal feature sequences?

Intention Monitor

School of Computing and Information Systems

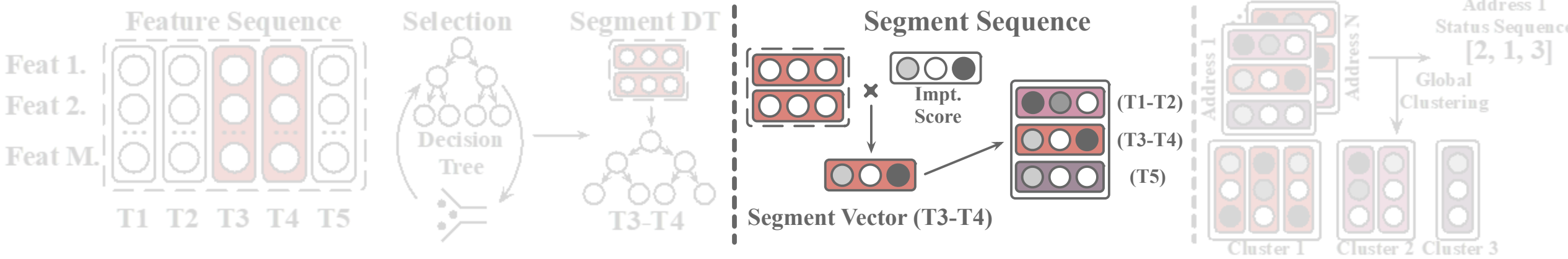


Step-1 Feature Selection & Segmentation

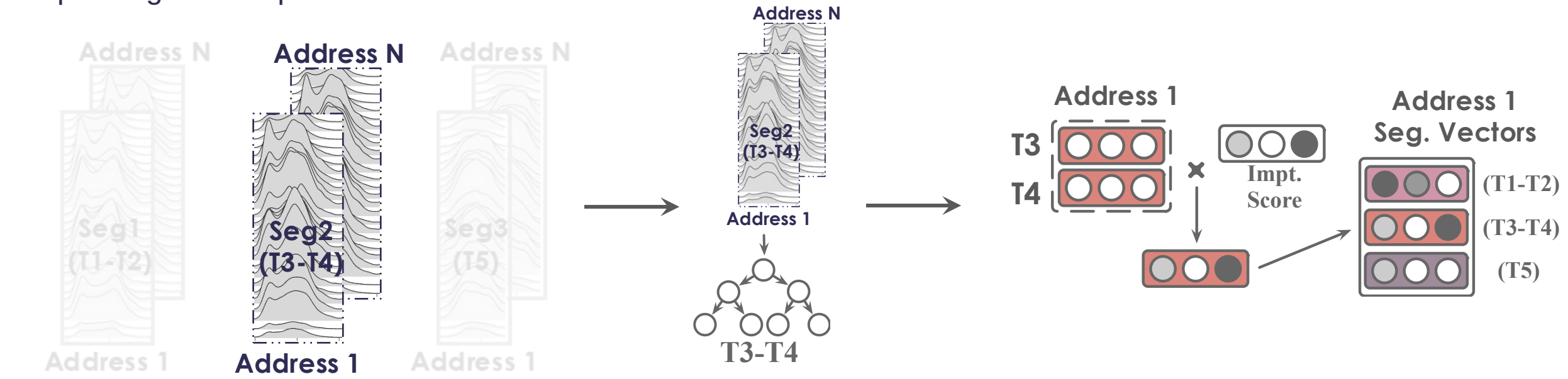


Intention Monitor

School of Computing and Information Systems

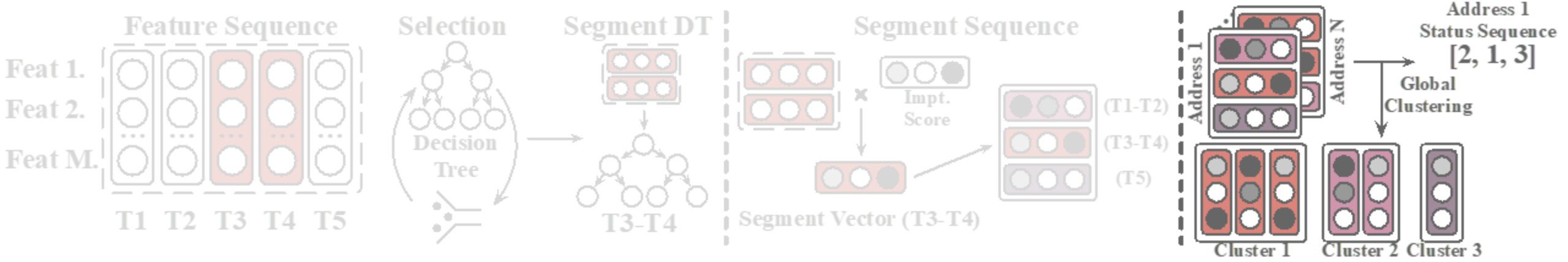


Step-2 Segment Representation

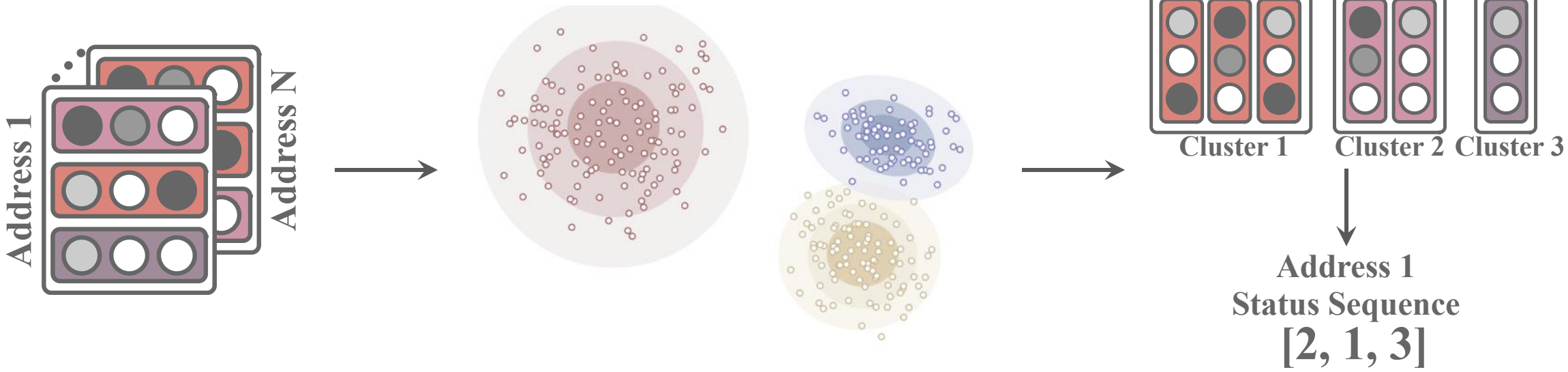


Intention Monitor

School of Computing and Information Systems

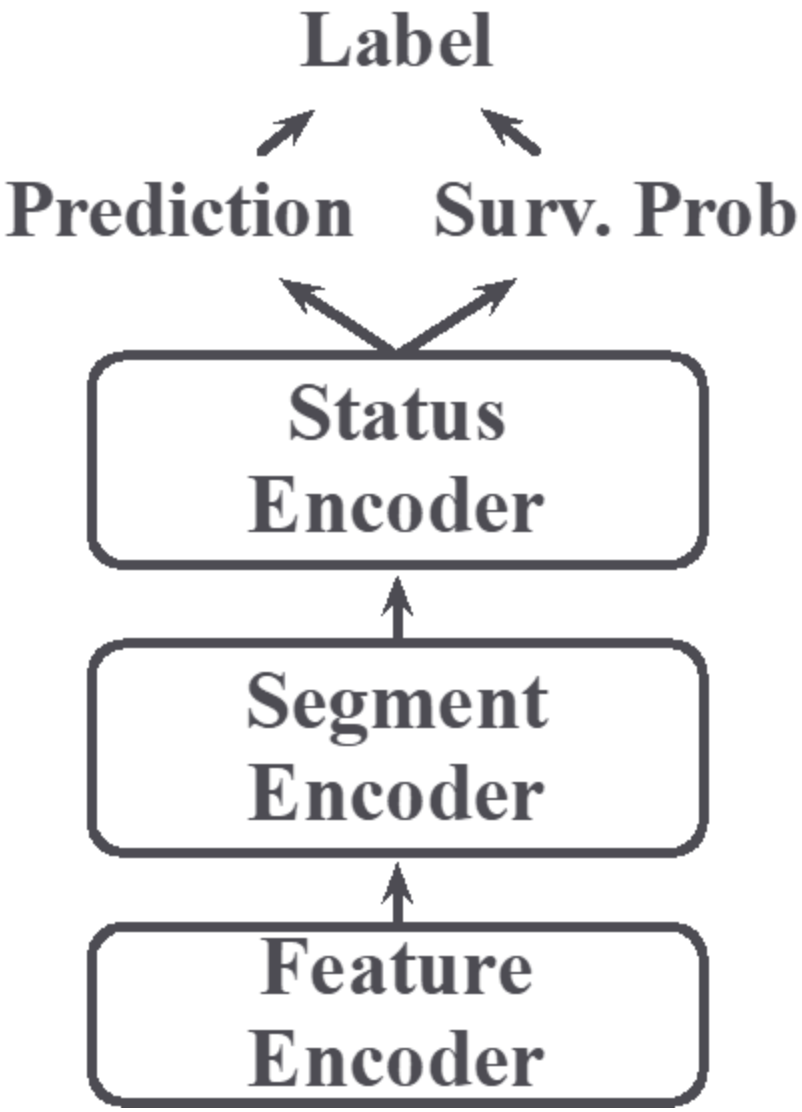


Step-3 Status Proposal



Prediction with Survival Analysis

Predictor

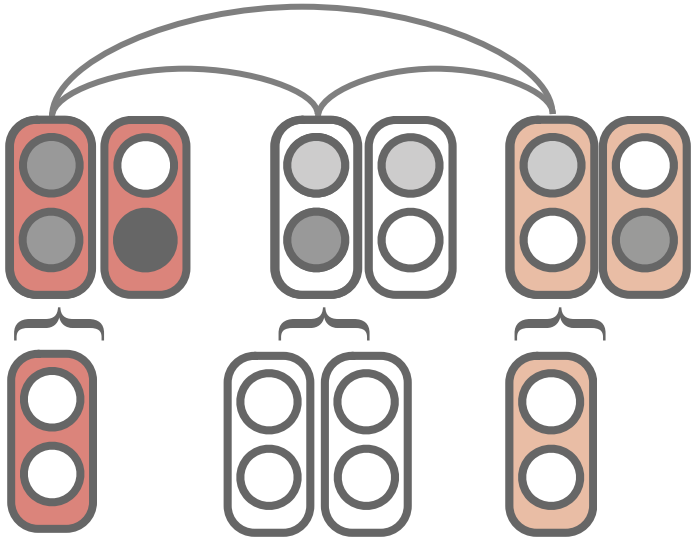


$$f^p = \sum_{i=b^p}^{e^p} \alpha_i f_i,\ \alpha_i = \exp(a_i) / \sum_{k=b^p}^{e^p} \exp(a_k),$$

$$a_i = W^a \tanh(W^{f,u}[f_i, u^p]),$$

$$F^p = \{\hat{f}^i\}_{i=1}^p = Concat(H_1^p, \cdots, H_h^p, \cdots, H_{N_h}^p) W^O,$$

$$H_h^p = Softmax(\frac{(QW_h^Q)(KW_h^K)^T}{\sqrt{d}}) V W_h^V,$$

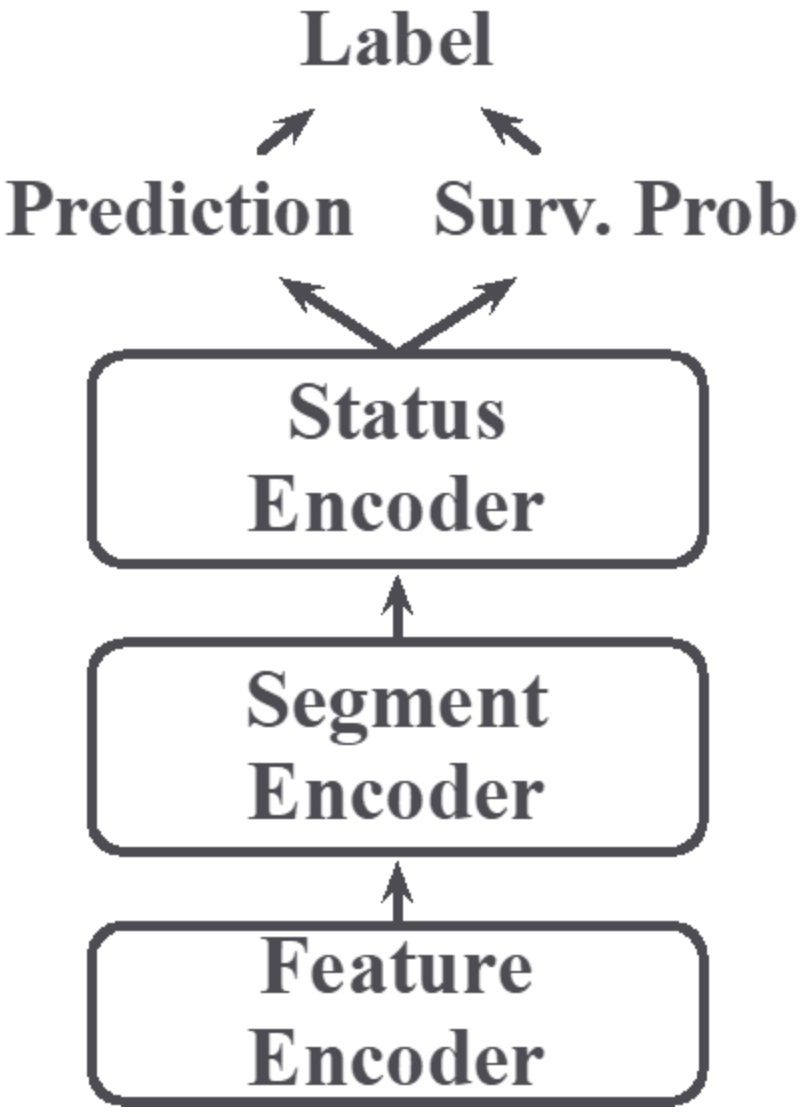


Feature Sequence (in hours)

$$\tilde{g}^i = W^g \tanh(W^{f,g}[g^i, \hat{f}^i]),$$

$$\tilde{u}^i = W^u \tanh(W^{g,u}[u^i, \hat{g}^i]),$$

Predictor



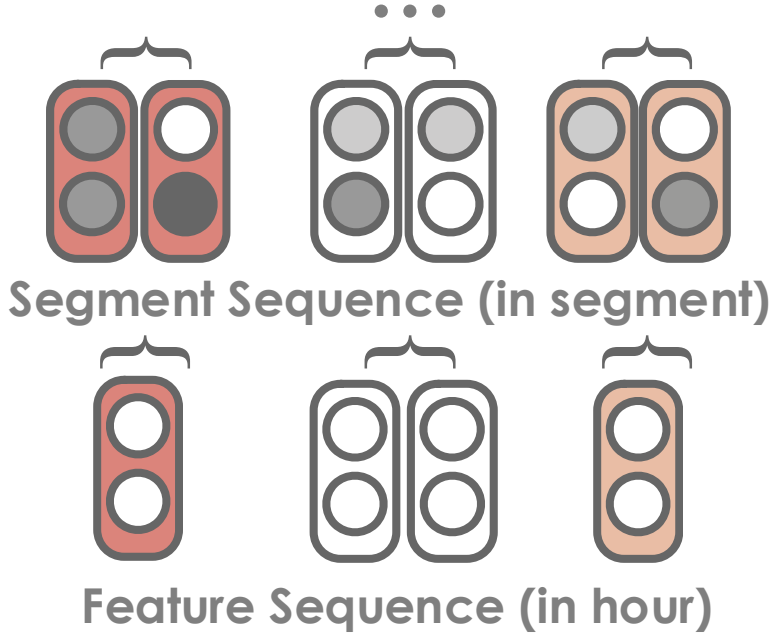
$$\hat{y}^t = S(t) * y^t + (1 - S(t)) * \hat{y}^{t-1}$$

$$y^p = \text{Sigmoid}(W^l * \bar{u}^p)$$

$$\lambda_t = ln(1 + \exp(W^{hz}\bar{u}^p)),$$

$$S(t) = exp(-\sum_{k=1}^{t} \lambda_k),$$

$$\bar{u}^p = Avg(\{\hat{u}^i\}_i^p)$$



Case Analysis

Case Analysis

School of Computing and Information Systems

Case Recap

Transaction e8b406091959700dbffc

ff30a60b190133721e5c39e89bb5fe23c5a554ab05ea

Txid	e8b406091959700dbffcff30a60b190133721e5c39e89bb5fe23c5a554ab05ea
Included in block	575013 (as a transaction number 138)
Time	2019-05-07 17:17:18
Sender	Binance.com
Fee	0.01188 BTC (99.15 satoshis/byte)
Size	11982 bytes

inputs: 71 (7074.19295031 BTC) unique addresses: 2, source transactions: 71

0.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ a98a74df...
1.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ dc03c5e9...
2.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ b3ca84de...
3.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 35f86114...
4.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ a3b14077...
5.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ d4aff83a...
6.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ af920705...
7.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 9afce068...
8.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ bd01d62c...
9.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ e50b5154...
10.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 1374e3cd...
11.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ d91ccfc1...
12.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ d42c2d3b...
13.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 156d3abe...
14.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 246b49ac...
15.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ dc22a158...
16.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 715f4cbd...
17.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 27c8f9e0...
18.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 8eef5bc2...
19.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 49c0b2d9...
20.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ fbcaa6f2...
21.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ bf45ad7b...
22.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 234e6c60...
23.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 08056916...
24.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ 032ffd2d...
25.	1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.	BTC	⇐ faa33c7e...

outputs: 44 (7074.18107031 BTC) unique addresses: 44, spent: 43 in 33 transactions

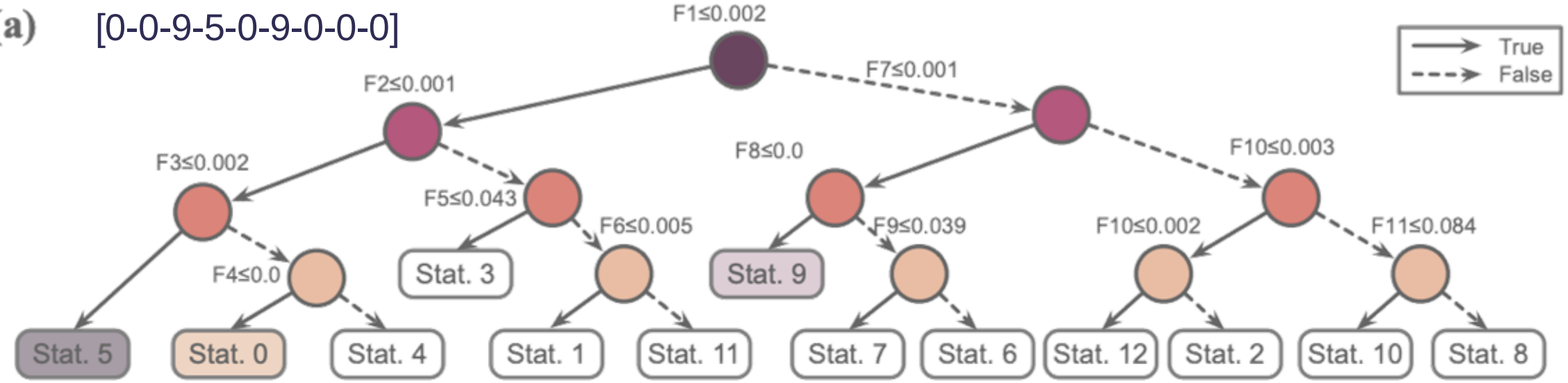
0.	bc1qp6k6tux6g3gr3sxw94g9tx4l0cjtu2pt65r6xp	[2e5ac3b67e]	555.997	BTC	6884775a... ⇒
1.	bc1qqp8pwq277d30cy7fjpvhcvhgztvs7v0nudgul5	[7f9e9afd92]	463.9975	BTC	8b1e6213... ⇒
2.	32LZ4wWwEhTzwtqAm2gPauktYZb5kQ6C5a	CoinPayments.net	0.0026	BTC	ccfe4342... ⇒
3.	3BMEXuoRza9EimRGSHGrwPmyFNUqWFpu8t	[0888b50bb7]	0.0746535	BTC	bf941a31... ⇒
4.	bc1qld27dqu6wrl4tmjdr8tl55qavmghwrr4ldh7qn	[7f9e9afd92]	473.9975	BTC	8b1e6213... ⇒
5.	3BMEXtMSkRt3wwXKytg7Nj86utJeSbwFHx	[51a9905c41]	0.17787495	BTC	be06bb29... ⇒
6.	bc1q8m9h3atn4cqeqhu3ekswdqxchp3g7d4v3qv3wm	[487907e868]	567.997	BTC	90ae2064... ⇒
7.	14QZ2wB8b8ZQNgb978Lwptdc8Vhv5aZQM2	[195be6cf37]	0.01944165	BTC	728f59a9... ⇒
8.	3L8JcsWNa3kuVaQJxAE1hhcoBT17rcJA6b	[00002dbb51]	0.01493527	BTC	db3e5299... ⇒
9.	bc1q7p6edvd4zvtya8uj366c23dan8pvlp503spucu	[66b7fc2922]	468.9975	BTC	bb0b41c2... ⇒
10.	bc1q93ecep2338dy9aauwyvh4g22t49rnedxl8z0tj	[589beb5a81]	0.1995	BTC	a8801564... ⇒
11.	bc1ql0wlnu80l8kctjzkzlzd72sdjqwuvruvgepceq	[7f9e9afd92]	383.998	BTC	8b1e6213... ⇒
12.	bc1q3ldtrr6xtpx8jam5gw68aaexz2wtluj0qullvr	[2377c0f10b]	189.999	BTC	7e615f3e... ⇒

Case Analysis

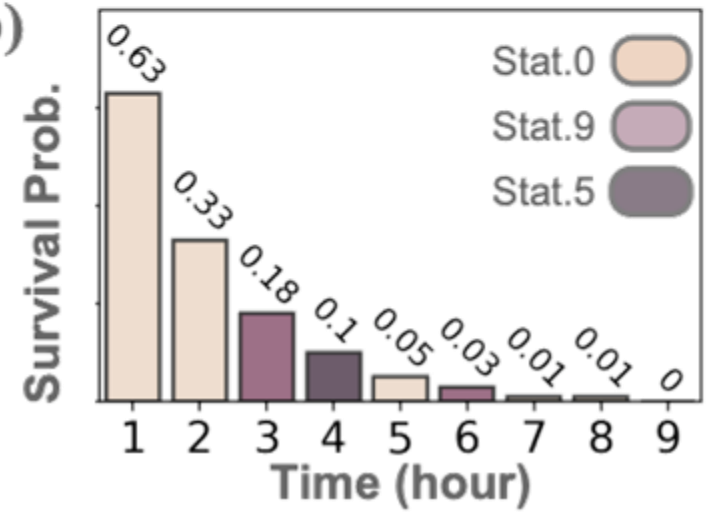
School of
Computing and
Information Systems

Sample Address Analysis

(a) [0-0-9-5-0-9-0-0-0]



(b)



(c)

F1: LT-BK-Max-Input-Num (Std)

F2: ST-BK-Min-Input-Num (Max)

F3: Life-time

F4: Spend-Tx-Num (Full-Time)

F5: ST-BK-Path-Num

F6: ST-BK-Hop-Length (Min)

F7: ST-BK-Min-output-Amt (Max)

F8: ST-BK-Hop-Length (Max)

F9: ST-FR-Height-Length (Std)

F10: LT-FR-Min-Trust (Std)

F11: LT-FR-Max-Input-Amt (Std)

Case Analysis

Sample Address Analysis

[0-0-9-5-0-9-0-0-0]

The Hack

Binance has been fairly forthcoming about the hack, detailing its impact in a blog post from Binance CEO Zhao Changpeng. "The hackers used a variety of techniques, including phishing, viruses and other attacks," wrote Zhao. "The hackers had the patience to wait, and execute well-orchestrated actions through multiple seemingly independent accounts at the most opportune time. The transaction is structured in a way that passed our existing security checks."

The hacker received 568 BTCs through 71 input TXs with no output.

↓

At the 13th hour, it received 0.00008642 BTC.

↓

At the 21st hour, it transferred out all its BTC.

Status 0

- Asset comes from a single source.
- No spend transaction.

Status 9

- The asset was obtained from a single source through a bunch of transitions.
- Each transition "peels" a certain amount off before passing it onto the receiver.

Status 5

- Still no spending transactions after the initial asset received from a single source at the early beginning.

Case Analysis

Based on the five transaction characteristics you provided, it is difficult to determine whether an address is classified as a "malicious" address without additional context or information.

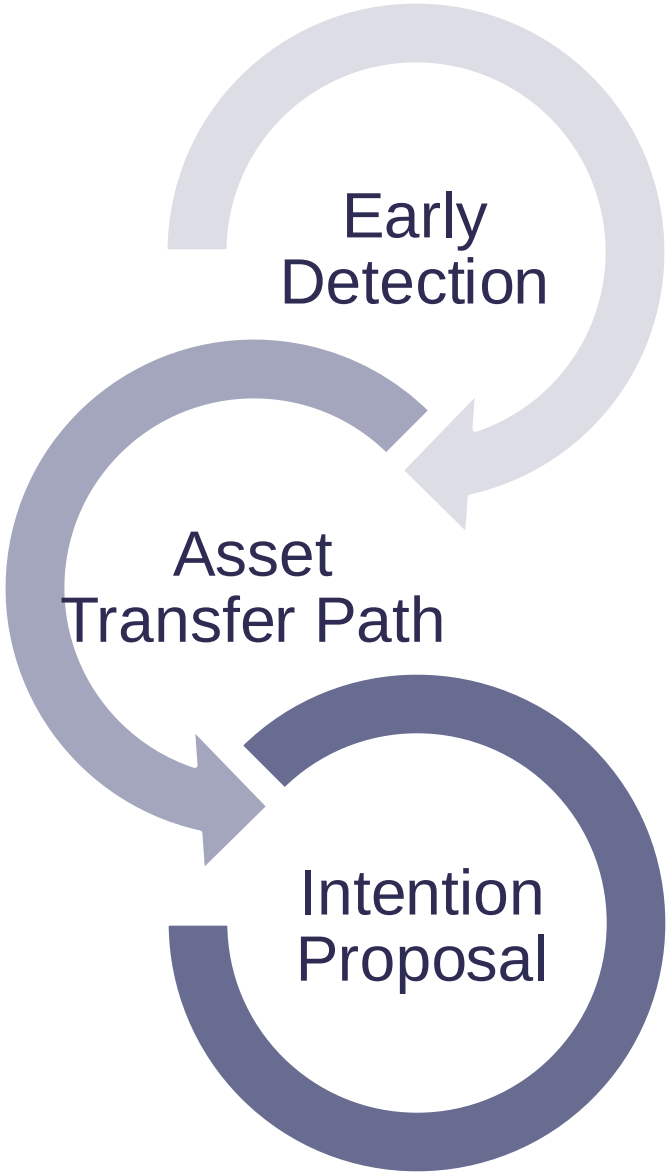
However, some possible reasons why an address with these characteristics could be considered malicious include:

1. The single source of the asset may be associated with illicit activity, such as theft or fraud.
2. The lack of spending transactions could indicate that the asset is being hoarded for future illicit purposes, such as money laundering or financing illegal activities.
3. The "peeling" of a certain amount off each transition could be a way to obfuscate the source of the funds and avoid detection by authorities.
4. The fact that the asset was obtained from a single source through a series of transitions may be an attempt to distance the funds from their original illicit source.

It's important to note that without more information or context, it's impossible to definitively classify an address as "malicious." However, the transaction characteristics you provided do raise some red flags and may warrant further investigation.

Conclusion

School of
Computing and
Information Systems



- Illicit early detection is necessary in BTC system.
- Asset flow gives more information at an early stage.
- Intention motifs can profile suspicious patterns.

Outline

- Introduction to Blockchain
- **Hands-on case 1**: Rat Trading Detection Example
- Traditional Methods on Crypto Crime Detection
- **Hands-on case 2**: NFT Wash Trading Detection Example
- Transfer Path-based Methods
- **Hands-on case 3**: Scam Detection Using Asset Transfer Path