Rensselaer

why not change the world?®

Combatting Crypto Scams on the Web with Semantic Graph Analysis

Dan Kazenoff, Oshani Seneviratne, Deborah L. McGuinness

Rensselaer Polytechnic Institute
dkazenoff@gmail.com, senevo@rpi.edu, dlm@cs.rpi.edu

IDEA

Rensselaer Institute for Data Exploration and Applications

Outline

1. Motivations
2. Recent Crypto Scams
3. Related Work
4. Integrative Blockchain Provenance Analyzer (IBPA)
    a. Design and Operation
    b. Results and Evaluation
5. Future Work

Motivation: Fix Rampant Misinformation on the Web about Cryptocurrencies

- Blockchain space is highly fragmented
- The space is especially highly susceptible to fraudulent activities
  - Deceptive smart-contracts
  - DeFi scams, & “pump and dumps”
  - “A-List” / “Airdrop” Phishing Scams
- Anonymity of many blockchain tokens act as a vessel for widespread phishing scams that so far are effectively untraceable
- **No technologies are widely available to identify crypto addresses involved in a phishing scam as “fraudulent”**

Rensselaer

why not change the world?®

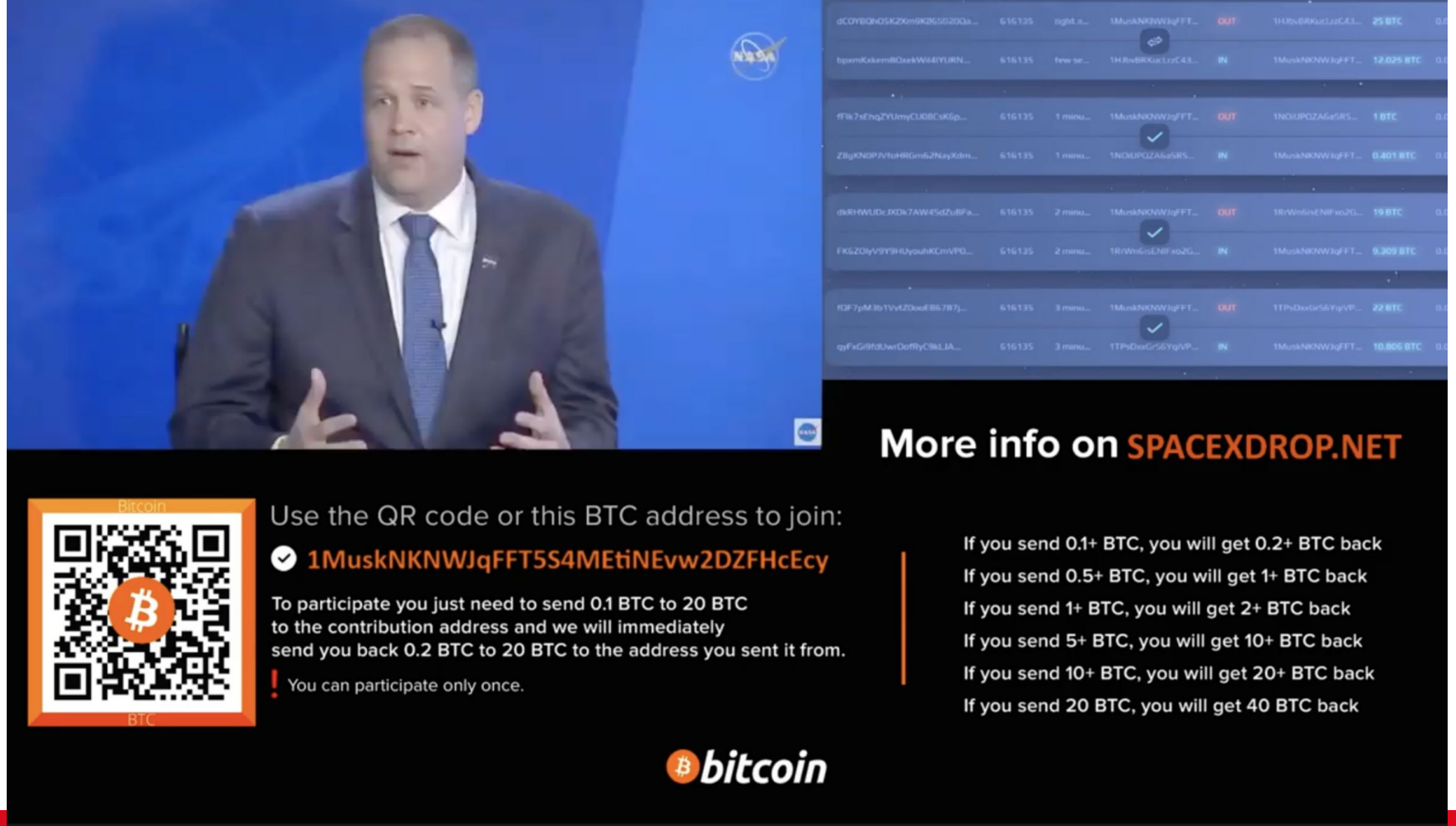
High-Profile

Crypto “Phishing” Scams

IDEA

Rensselaer Institute for Data Exploration and Applications


Tweet
Joe Biden
@JoeBiden
I am giving back to the community.
All Bitcoin sent to the address below will be sent back doubled! If you send $1,000, I will send back $2,000. Only doing this for 30 minutes.
bc1qxy2kgdygjrsqtzq2n0yrf2493
Enjoy!
4:22 PM · 15 Jul 20 · Twitter Web App
2,375 Retweets and comments
1,739 Likes



#NASA #SpaceX #Live

Falcon 9 and Crew Dragon launch | Starlink Mission | Elon Musk | SpaceX Conference | Live

WELCOME TO 5000 BTC GIWEAVAY

SPACEX

BACK BONUS SYSTEM

IF YOU SEND 0.1+ BTC, YOU WILL BE AIRDROPPED 0.2+ BTC BACK

IF YOU SEND 0.5+ BTC, YOU WILL BE AIRDROPPED 1+ BTC BACK +10% BONUS

IF YOU SEND 1+ BTC, YOU WILL BE AIRDROPPED 2+ BTC BACK +25% BONUS

IF YOU SEND 5+ BTC, YOU WILL BE AIRDROPPED 10+ BTC BACK +40% BONUS

IF YOU SEND 10+ BTC, YOU WILL BE AIRDROPPED 20+ BTC BACK +60% BONUS

5000 BTC GIVEAWAY

TO PARTICIPATE YOU JUST NEED TO SEND BEETWEEN 0.1+ BTC TO 10+ BTC TO THE CONTRIBUTION ADDRESS WE WILL IMMEDIATELY SEND YOU BACK BEETWEEN 0.2+ BTC TO 20+ BTC TO THE ADDRESS YOU SEND IT FROM. EVERY PERSON CAN PARTISIPATE ONLY ONE TIME.

Bitcoin

BTC

1xGdMFTpCkQJWiuZSufa8MKrsrrgbPyxQ

Just Log into your mobile APP;
Scan QR Code!
Send amount of 0.1+ BTC to 20+ BTC to participate.
REMEMBER! You can participate only once!

MORE INFO: **MUSKXEVENT.SITE**

bitcoin

#NASA #SpaceX #Live

Falcon 9 and Crew Dragon launch | Starlink Mission | Elon Musk | SpaceX Conference | Live

2,454 watching now • Started streaming 57 minutes ago

286 10 SHARE SAVE

bitcoin

Live

STEVE WOZNIAK

The Past, Present and Future of Crypto

5000 BTC GIVEAWAY

More info on WOZBTC[redacted]

Use the QR code or this BTC address to join:

1WoZp2r6NMXbqKk7nYm6WUpWP[redacted]

To participate you just need to send 0.1 BTC to 20 BTC to the contribution address and we will immediately send you back 0.2 BTC to 40 BTC to the address you sent it from.

! You can participate only once.

- If you send 0.1+ BTC, you will get 0.2+ BTC back
- If you send 0.5+ BTC, you will get 1+ BTC back
- If you send 1+ BTC, you will get 2+ BTC back
- If you send 5+ BTC, you will get 10+ BTC back
- If you send 10+ BTC, you will get 20+ BTC back
- If you send 20 BTC, you will get 40 BTC back

bitcoin


Steve Wozniak interview: Blockchain technology, AI, Crypto, Bitcoin BTC Halving 2020
Live
STEVE WOZNIAK
The Past, Present and Future of Crypto
5000 BTC
GIVEAWAY
Use the QR code or this BTC address to join:
1WoznScUTGbqGTqqeLPBgAVPgipx64URX
To participate you just need to send 0.1 BTC to 20 BTC
to the contribution adress and we will immediately
send you back 0.2 BTC to 20 BTC to the adress you sent it from
More info on WozDrop.me
If you send 0.1+ BTC, you will get 0.2+ BTC back
If you send 0,5+ BTC, you will get 1+ BTC back
If you send 1+ BTC, you will get 2+ BTC back
If you send 5+ BTC, you will get 10+ BTC back

Can we figure out a way to flag these scam addresses (and addresses likely associated with them) before they offload their wallets at an exchange?

Rensselaer

why not change the world?®

Selected Related Work

IDEA

Rensselaer **Institute for Data Exploration and Applications**

Phillips & Wilder: Clustering Replicated Phishing Sites

- Searched for similarities between operators of various, seemingly disconnected scam operations
- Discovered many operations were in fact operated by the same campaigns via clustering
- Stressed need for exchanges to be more vigilant

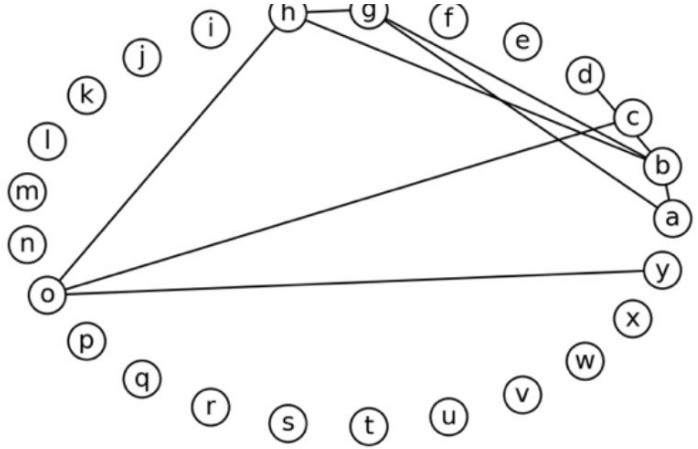


Fig. 5. Blockchain cluster overlap between different campaigns

Bartoletti et al.: Data Mining for Detecting BTC Ponzis

- Fraud detection to cybercrime analysis in BTC is an unexplored field
- Translate data mining techniques from credit card operations to Bitcoin Ponzi Schemes
- Created public dataset for systematic evaluations + comparisons of scams
- Automatic analysis comparing various ML approaches (supervised)
- Unclear how to operate for identifying addresses in real-time

Rensselaer

why not change the world?®

IBPA:

Integrative Blockchain Provenance Analyzer

IDEA

Rensselaer **Institute for Data Exploration and Applications**

Goal: Derive Meaning from Raw Ledger Data

Latest Transactions

Hash	Time	Amount (BTC)	Amount (USD)
aad1cf51f0c90f8083ab04...	01:44	0.01126765 BTC	$150.21
b1eba0ee41d83254317a6...	01:44	0.20907099 BTC	$2,787.13
4a84b753f8cc22190718b...	01:44	2.23379432 BTC	$29,778.78
1903d996c52c3d164fbb4...	01:44	0.04560389 BTC	$607.95
c206f9c030f7ceb33c070...	01:44	0.06598789 BTC	$879.69
88092a2709986ffcee42c...	01:44	0.00761768 BTC	$101.55

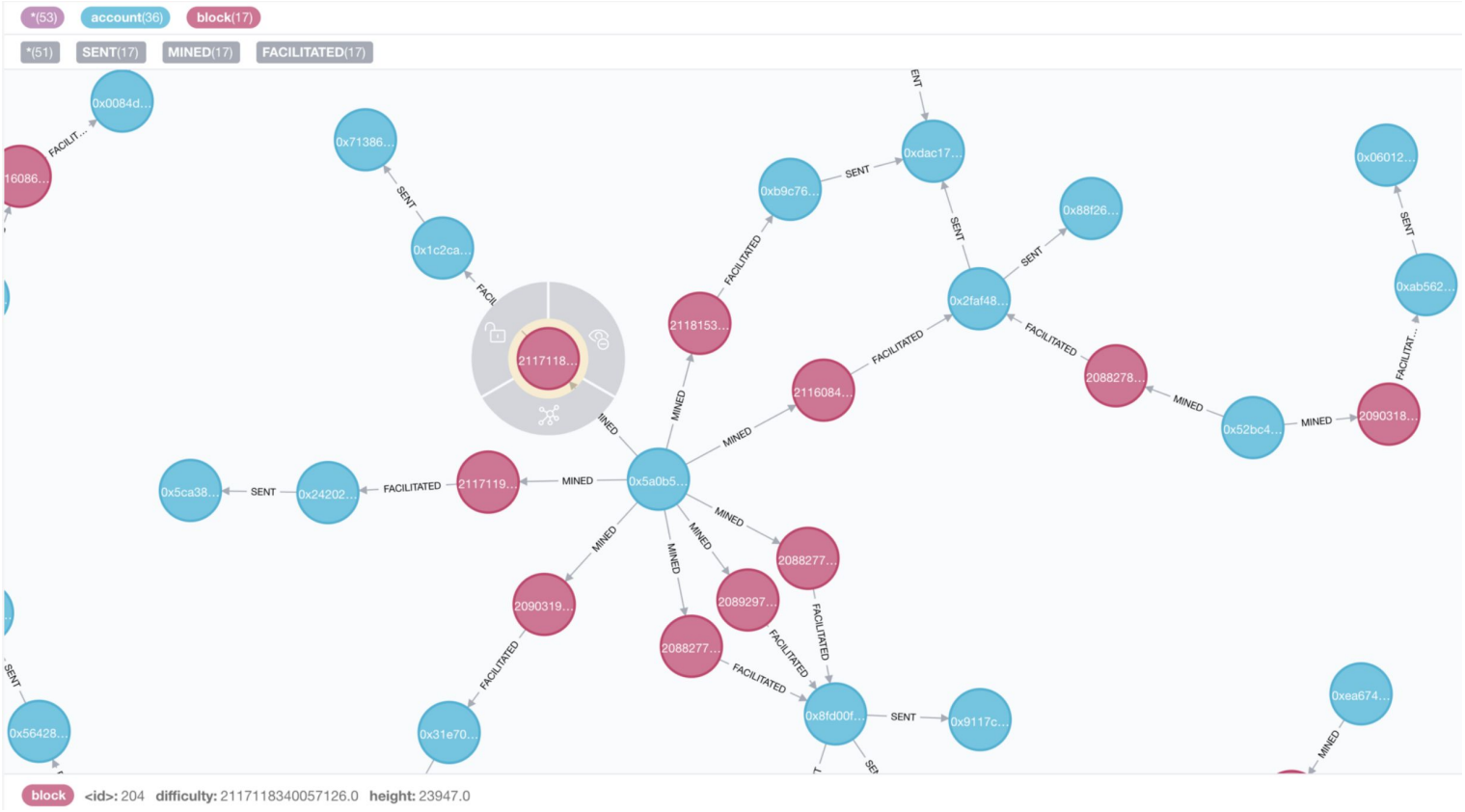
https://www.blockchain.com/explorer

Latest Transactions

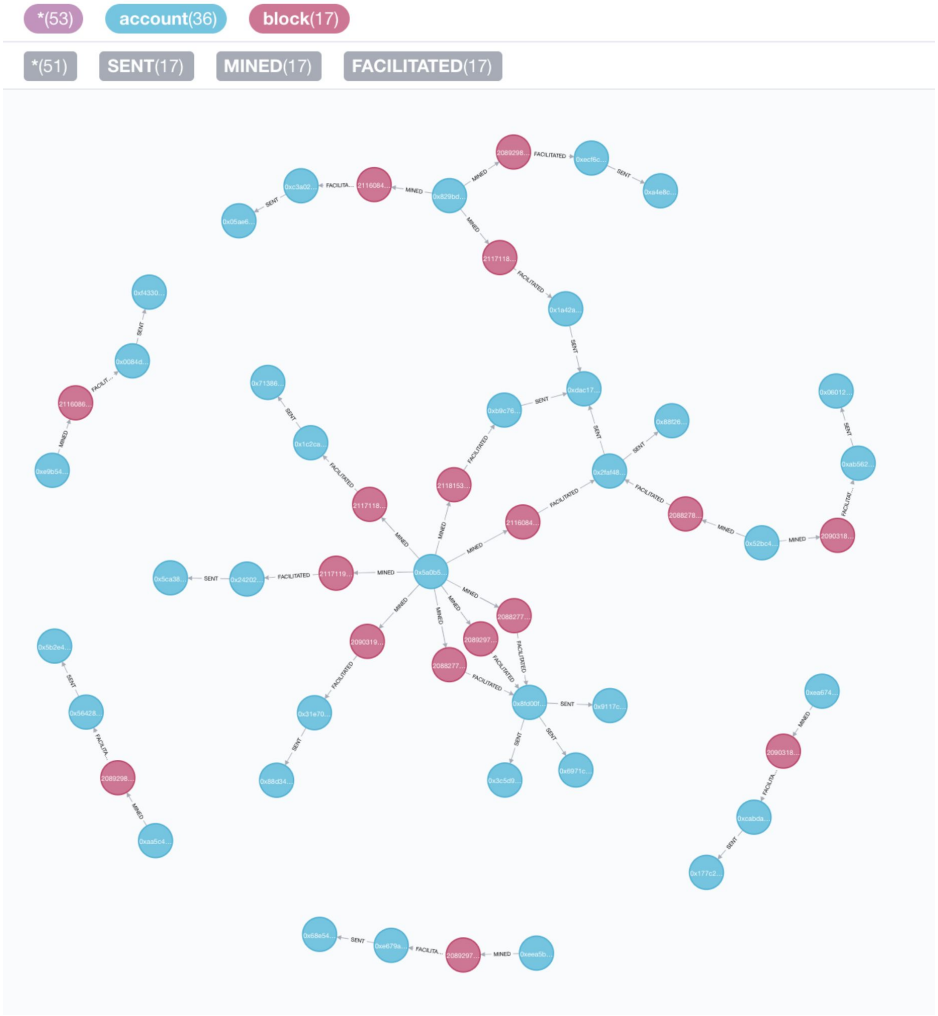
Tx	0x39fc91ef7d1e... 22 secs ago	From 0xee576686b08ffbaaf8... To 0x96f9632b25f874769...	0 Eth
Tx	0x6d2e354ab3e... 22 secs ago	From 0xc4fffd98fb3caec63cc... To 0x983cdcb1fbb14692e...	0.003 Eth
Tx	0x9859f26bd57... 22 secs ago	From 0xc63fc6907d3f9657a... To 0x1457b31c41ae3a13d...	0.059 Eth
Tx	0x5dbad46205c... 22 secs ago	From 0x098306de8c7701999... To 0x5fda1bb20918e5797...	0 Eth
	0xce76b73caec...	From 0x9983cfd319085687b...	0 Eth

https://etherscan.io

Can We Use Provenance to Identify Suspicious Addresses?

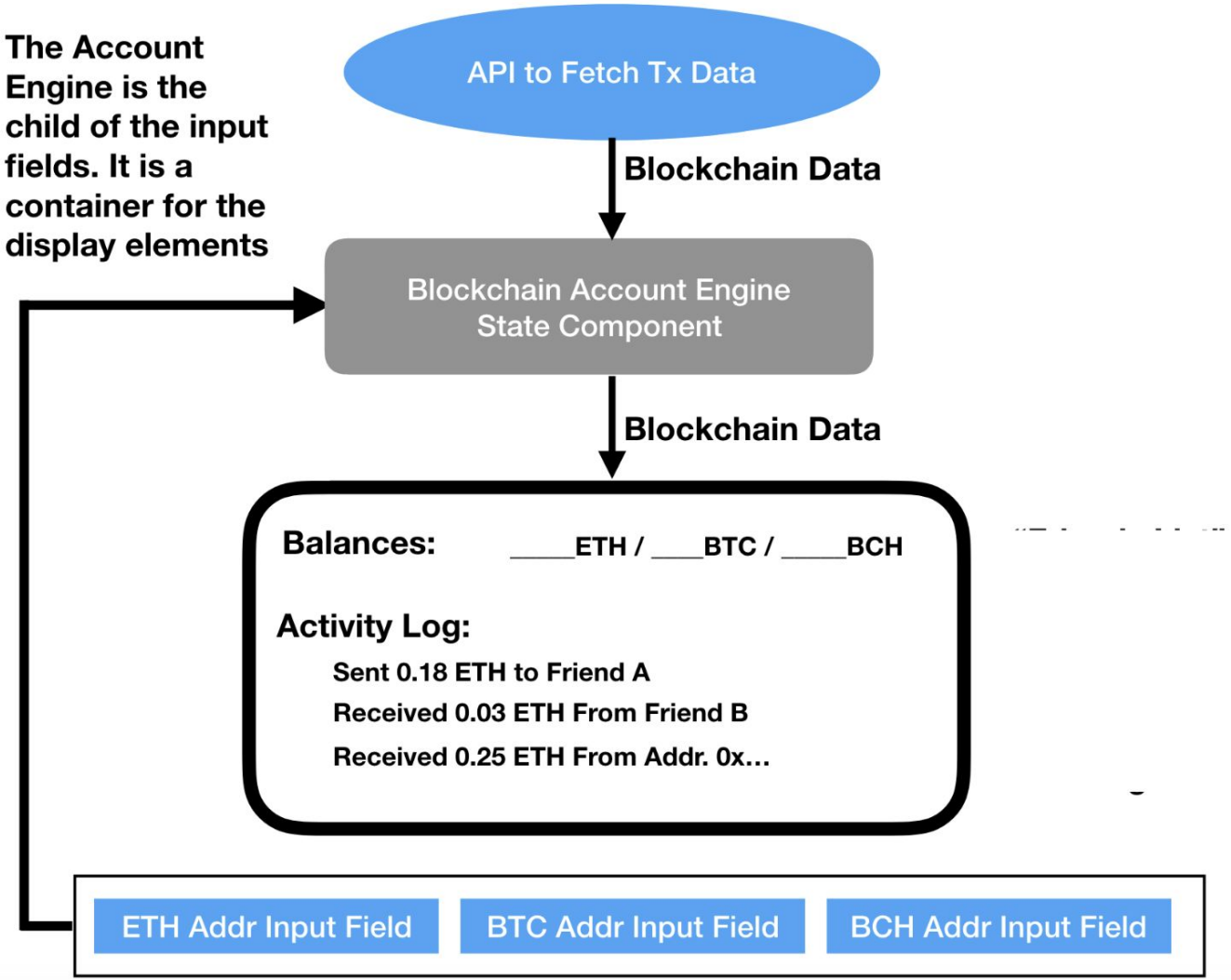


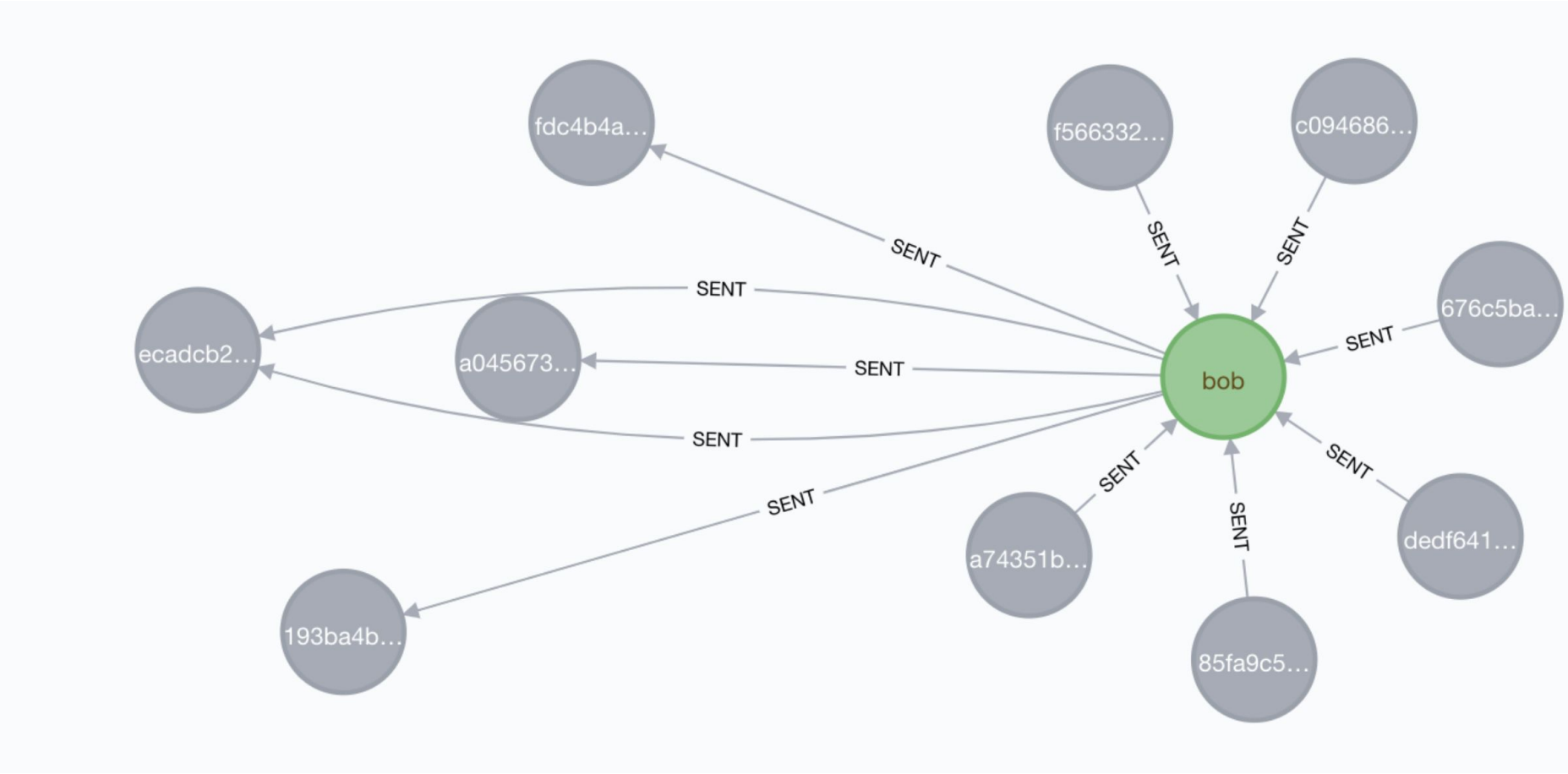
Answer:
Yes we can



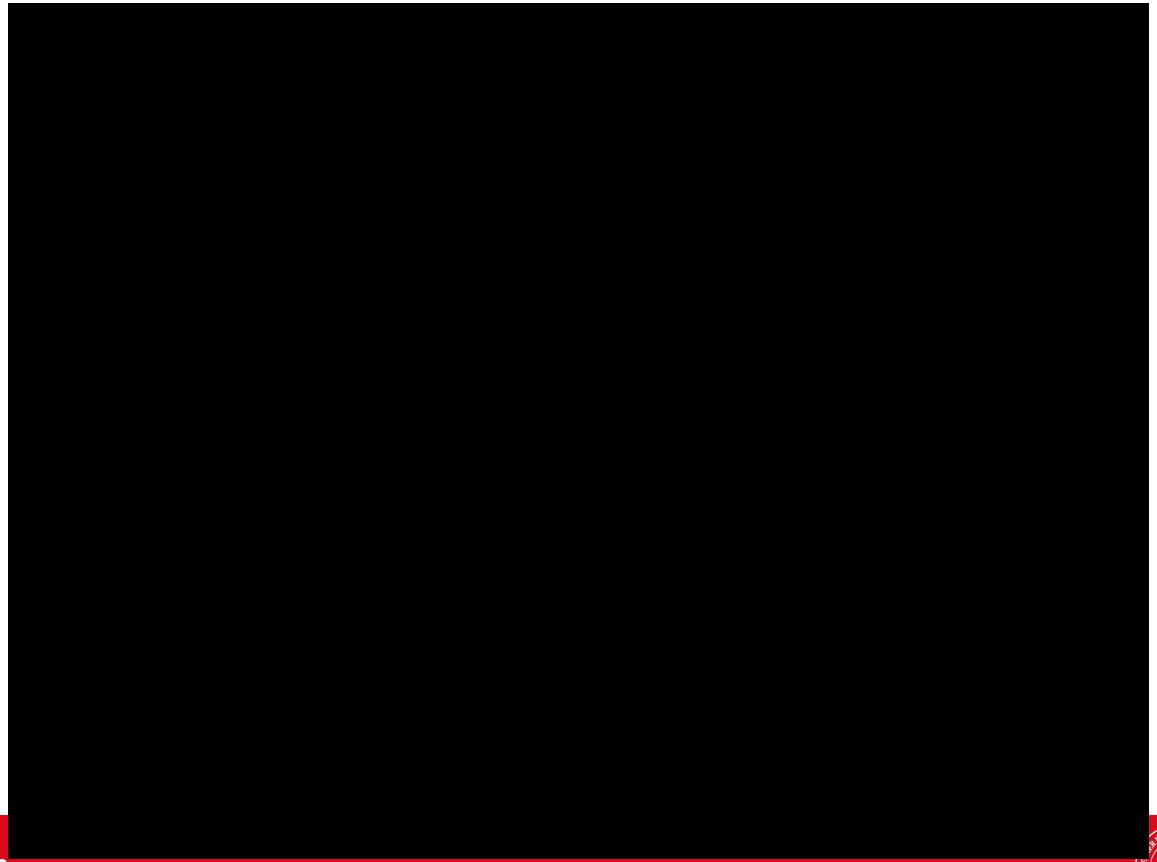
Design of the IBPA: Extract Transform and Load

- ***E**xtracts* raw tx data from a given blockchain network via API calls, initiated by the user on the frontend
- ***T**ransforms* that data, filtering the metrics / keys that matter (what were the final input and output addresses + amounts to the transaction)
- ***L**oads* that data into a Neo4j database for further querying by the IBPA
- Analysis: Runs tests using CypherQL+JS to return a predicted answer

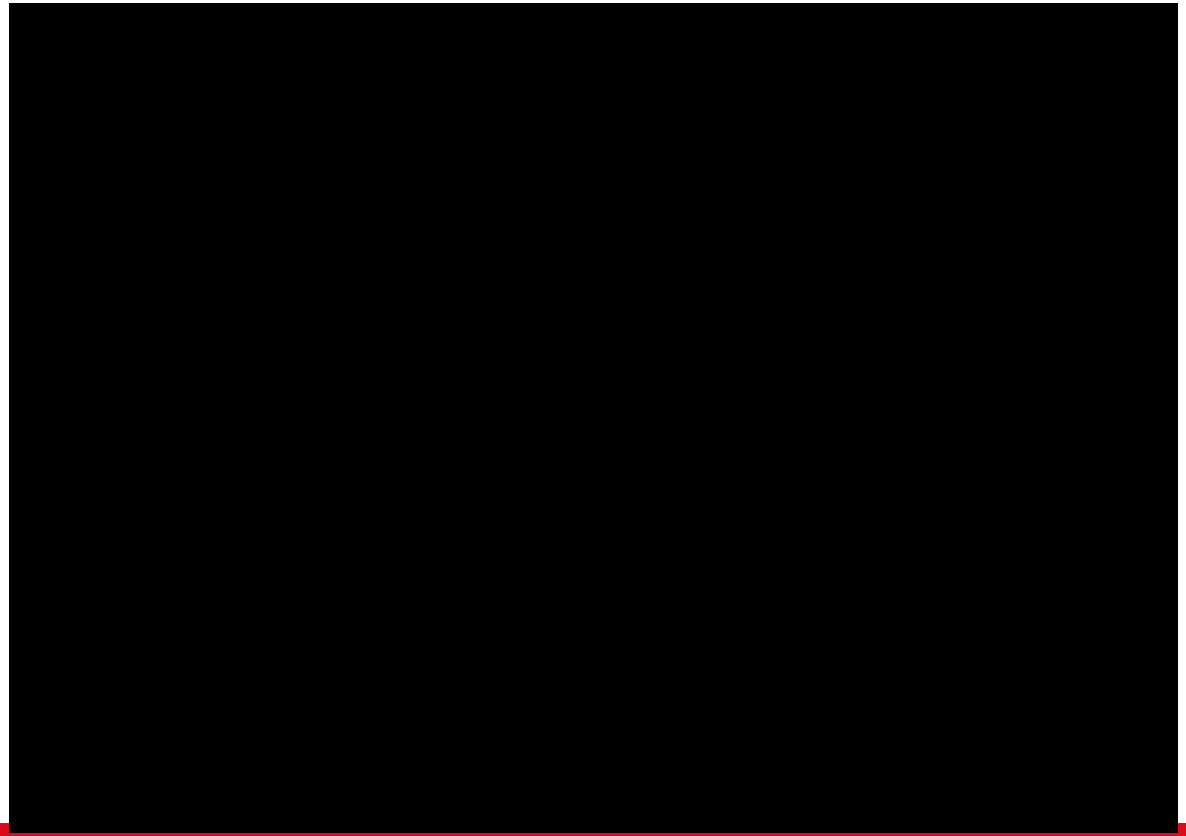

The Account Engine is the child of the input fields. It is a container for the display elements
API to Fetch Tx Data
Blockchain Data
Blockchain Account Engine State Component
Blockchain Data
Balances: ____ETH / ____BTC / ____BCH
Activity Log:
Sent 0.18 ETH to Friend A
Received 0.03 ETH From Friend B
Received 0.25 ETH From Addr. 0x...
ETH Addr Input Field
BTC Addr Input Field
BCH Addr Input Field


fdc4b4a…
f566332…
c094686…
ecadcb2…
a045673…
bob
676c5ba…
a74351b…
dedf641…
193ba4b…
85fa9c5…
SENT

IBPA Operation: Scam Addresses



IBPA Operation: Control Addresses



IBPA Results:

Avg Tx/Hr	Median Tokens	Incoming Nodes	Outgoing Nodes	Incoming Rels	Outgoing Rels	IBPA Score & Classification
0.015	0.0133	7	6	7	6	1 (regular)
0.015	0.6035	5	5	5	5	1 (regular)
<0.01	0.0003	9	0	9	0	1 (regular)
<0.01	0.7922	1	1	1	1	0 (regular)
<0.01	0	0	5	5	5	1 (regular)
0.07	0.01	7	2	13	11	5 (scam)
3.00	0.01	7	3	6	5	9 (scam)
8.73	0.01	11	1	11	5	11 (scam)
0.41	0.005	6	1	6	1	7 (scam)
4.00	0.022	6	3	11	24	3 (scam)

Table 1: IBPA results for regular (non-scam) and scam accounts. Any score >= 3 results in a failing score and the corresponding address is flagged as a suspicious address meeting fraudulent criteria.

IBPA Evaluation

- Does an effective job at quickly calculating a reasonable suspicion flag for a selected address
- Creates an easy-to-query graph representation of the surrounding nodes of interest
- Can be easily modified to quickly evaluate a large cluster of addresses
- Is a necessary first step in logically tackling identification of fraudulent actors

Rensselaer

why not change the world?®

Future Work

IDEA

Expanding the System

- Comprehensive data extraction pipeline
  - Scour the web intelligently for similar scams to plug into the evaluator
  - Combine with other semantically annotated data on the Web for enhanced provenance analysis
- Building out for additional network use-cases
  - Works for standard L1's, but what about L2 with state channels?
  - What about other privacy-enabled L1's?
- Large-scale testing
- Customize for exchange use / Metamask integration?

Summary

- There are many cryptocurrency scams on the Web
- We need tools and techniques to identify whether the addresses advertised are scam addresses or not
- We utilized network analysis techniques to identify scam addresses with good accuracy
- We have an initial encoding of the transaction graphs represented primarily using the provenance ontology
- We are expanding this work to include other information available on the “Semantic Web”

Questions/Comments?

Dan (dkazenoff@gmail.com), **Oshani** (senevo@rpi.edu), and **Deborah** (dlm@cs.rpi.edu)

Work partially funded from Tetherless World Constellation at RPI (dlm@cs.rpi.edu). and IBM Research AI through the AI Horizons Network.