Blockchainvis suite

A set of tools for bitcoin blockchain analysis and forensics

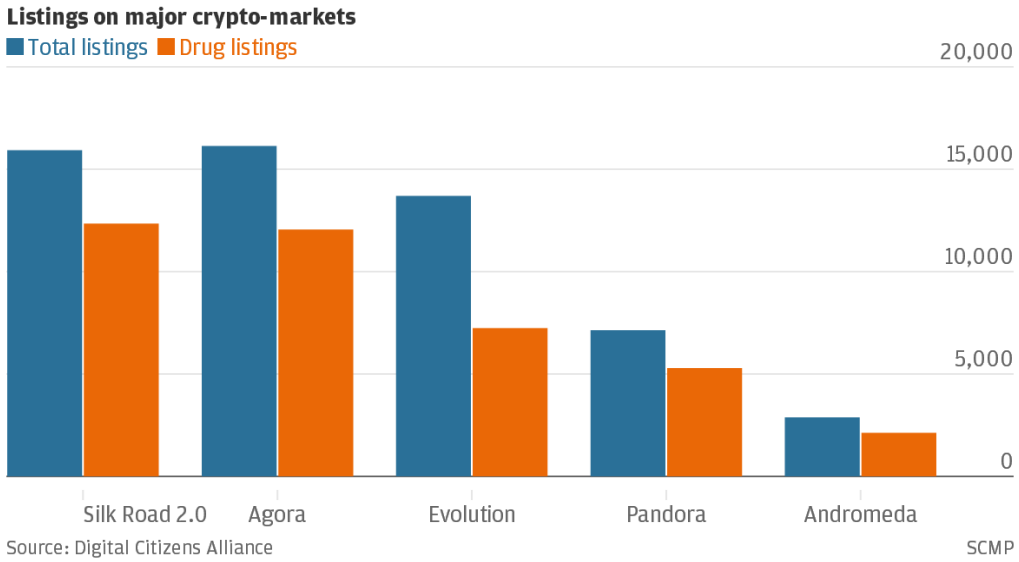
Stefano Bistarelli, Francesco Santini

With the help of

Ivan Mercanti, Francesco Moca, Eugenio Paluello, Matteo Parroccini, Emanuele Procacci, Dan Rusnac, Alessio Santoru,

Introduzione

I metodi di pagamento anonimi vengono ampiamente utilizzati per **scopi illeciti**: droga, armi o documenti falsi sono solo alcuni dei “prodotti” che è possibile acquistare online con una moneta anonima.



Confronto tra il numero di annunci totali e quelli di droga nei più famosi cripto-mercati

Ransomware

Riscatto da pagare con un metodo di pagamento anonimo: il **bitcoin**

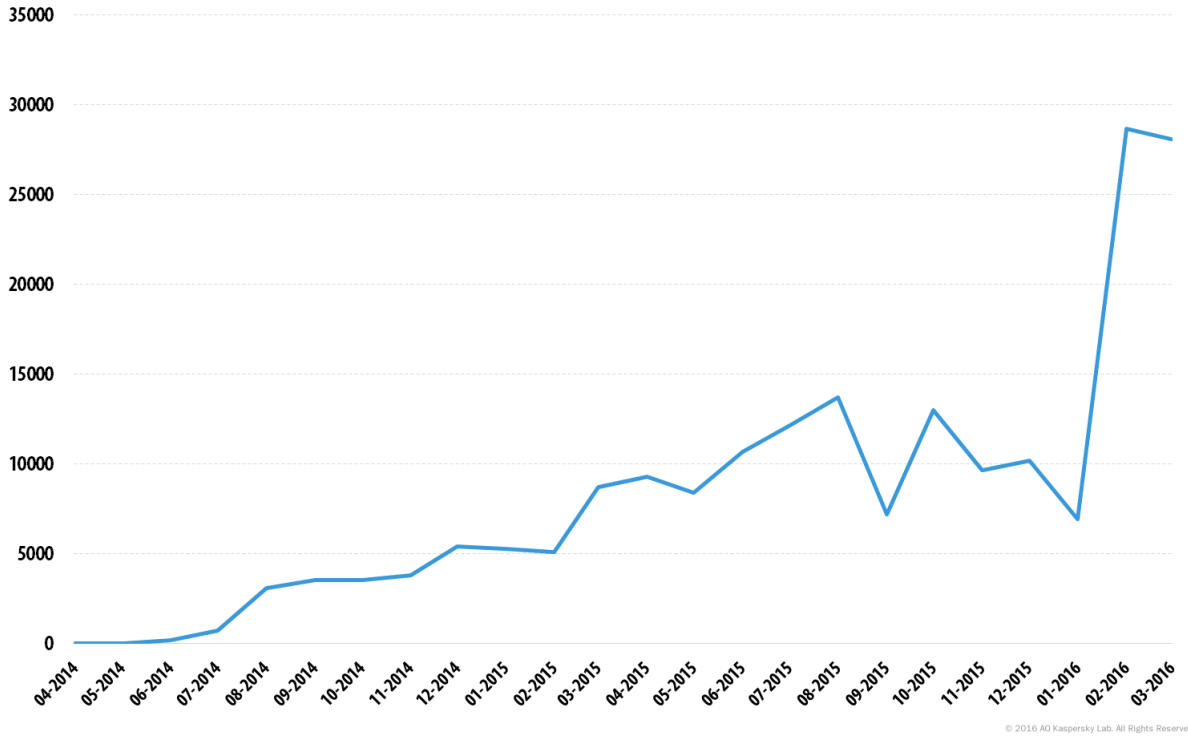


Grafico che mostra il numero di utenti colpiti da un ransomware in un dispositivo mobile (Kaspersky)

Bitcoin: Criptovaluta e Protocollo

Rete bitcoin formata da:

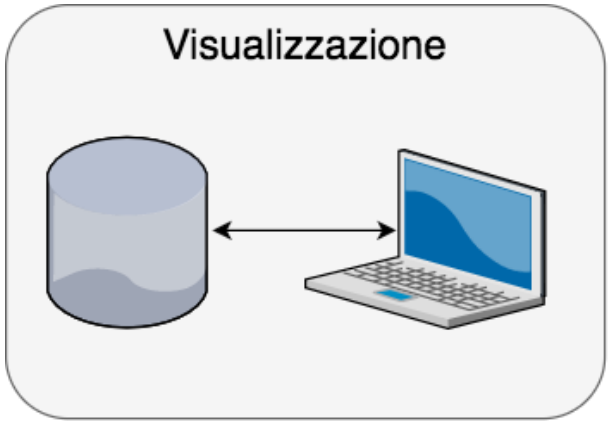
- Blockchain
- Transazioni
- Indirizzi Bitcoin

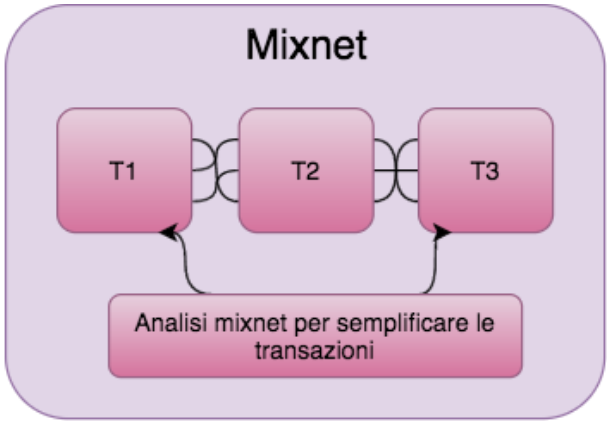
Identificare i pagamenti anonimi

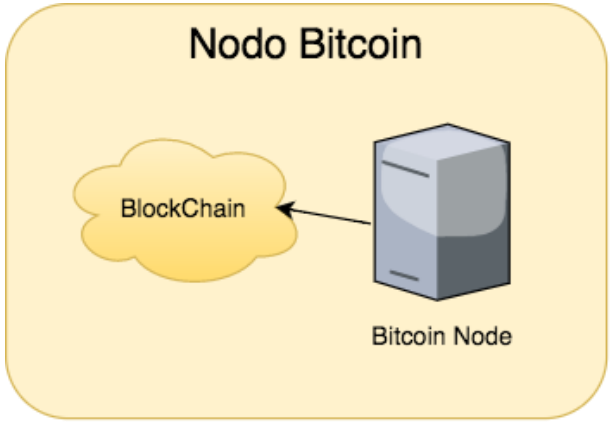
Necessità di:

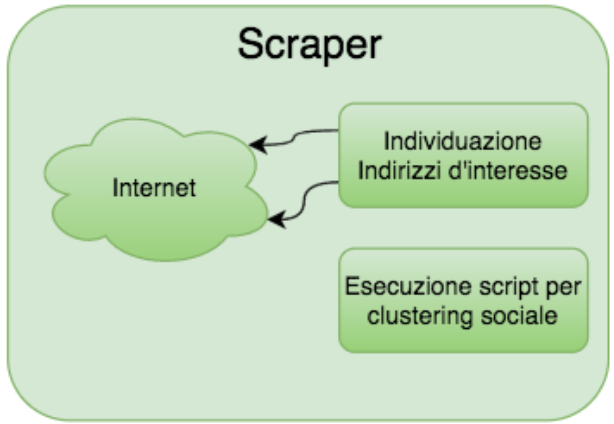
- Recuperare informazioni importanti per l’analisi
- Interrogare le risorse pubbliche disponibili
- Raggruppare sotto un unico utente più informazioni
- Visualizzare le informazioni e i dati ad esse correlate

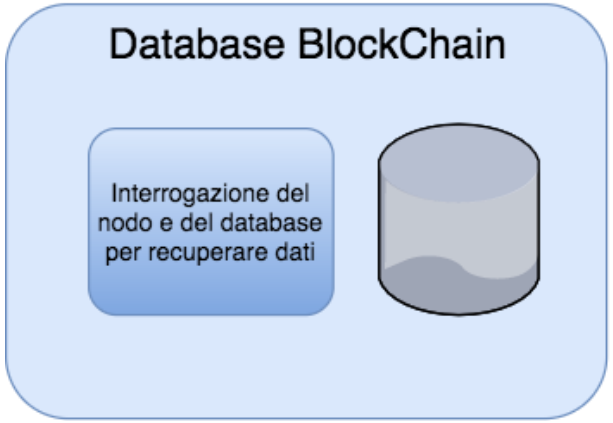
Un framework modulare

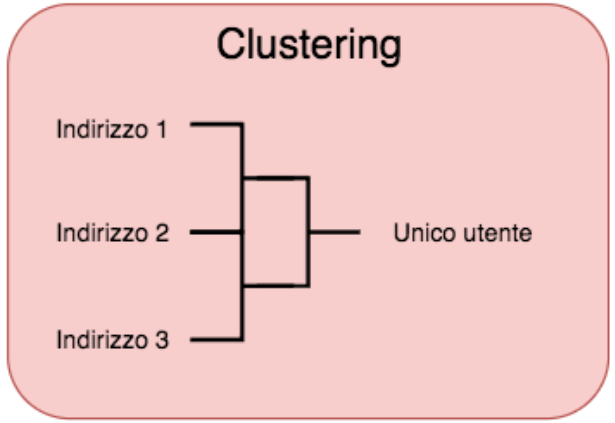




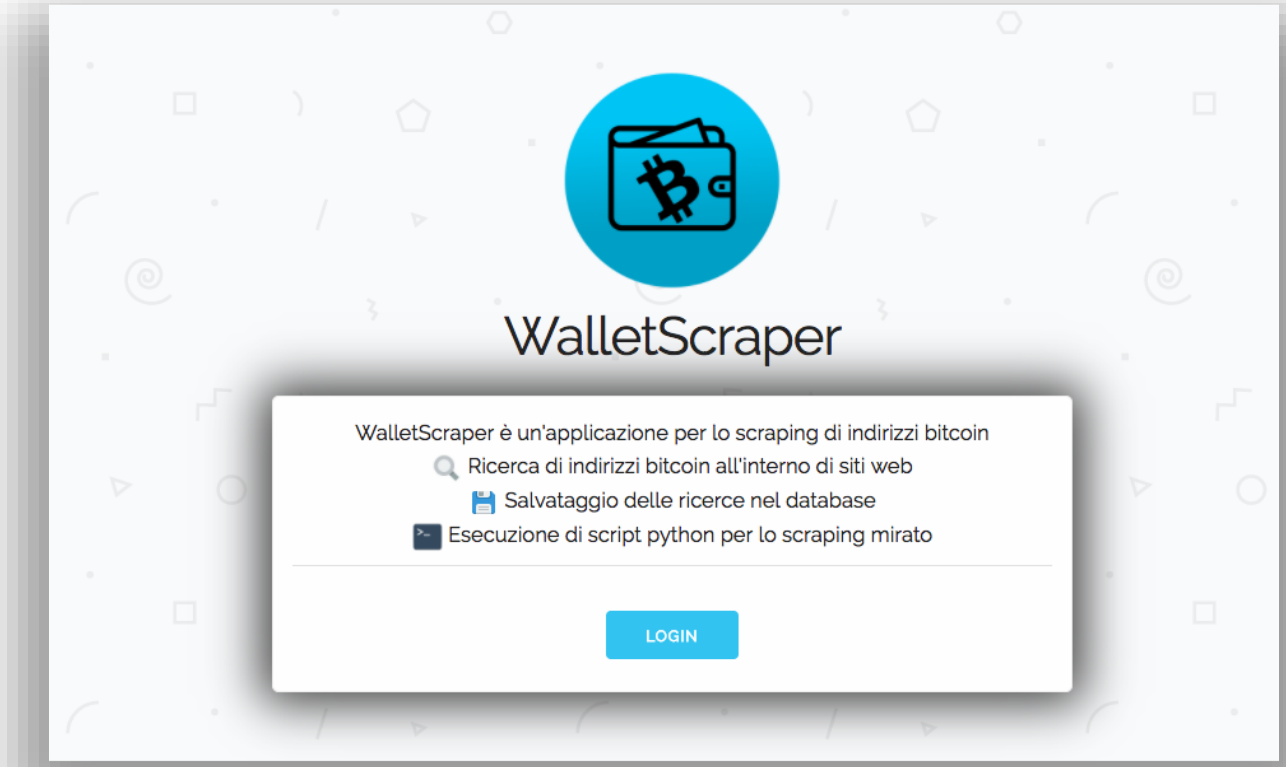








Un framework modulare: Scraper



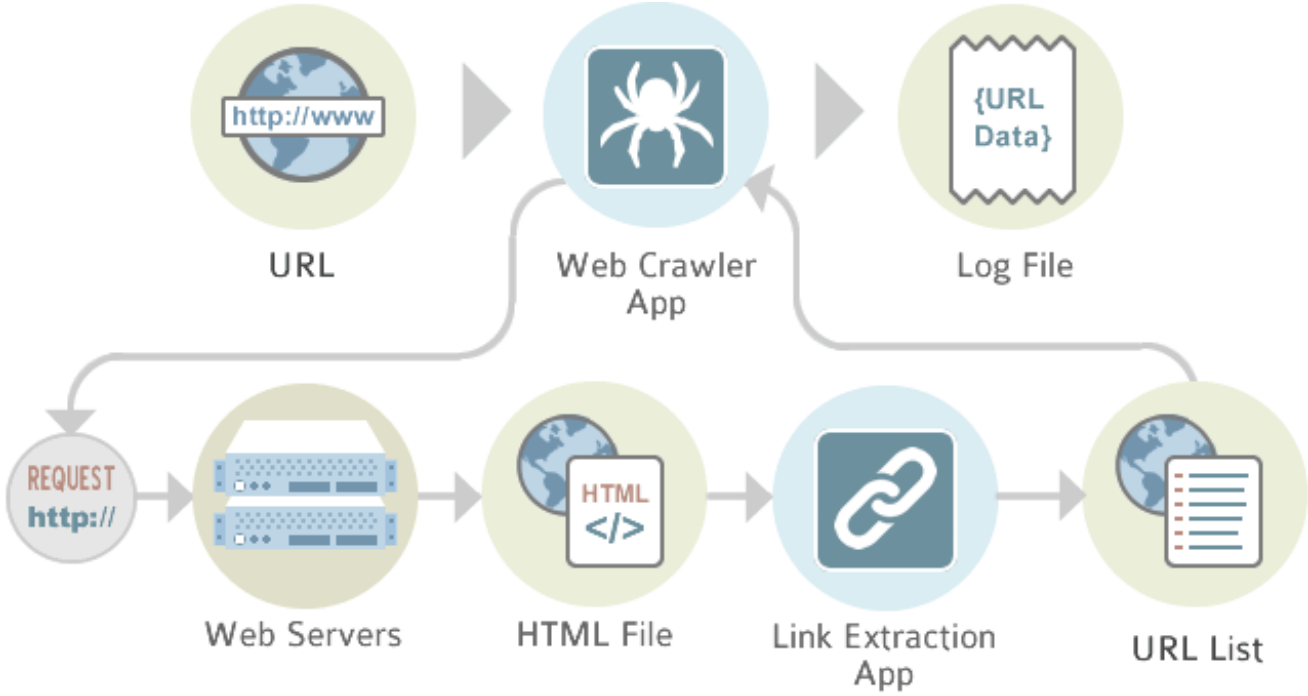
Indirizzi bitcoin

- Stringhe lunghe da 26 a 35 caratteri
- Primo carattere: 1 o 3
- Non contengono caratteri ambigui (O 0 I l) nella codifica utilizzata (base58)

Esempio:

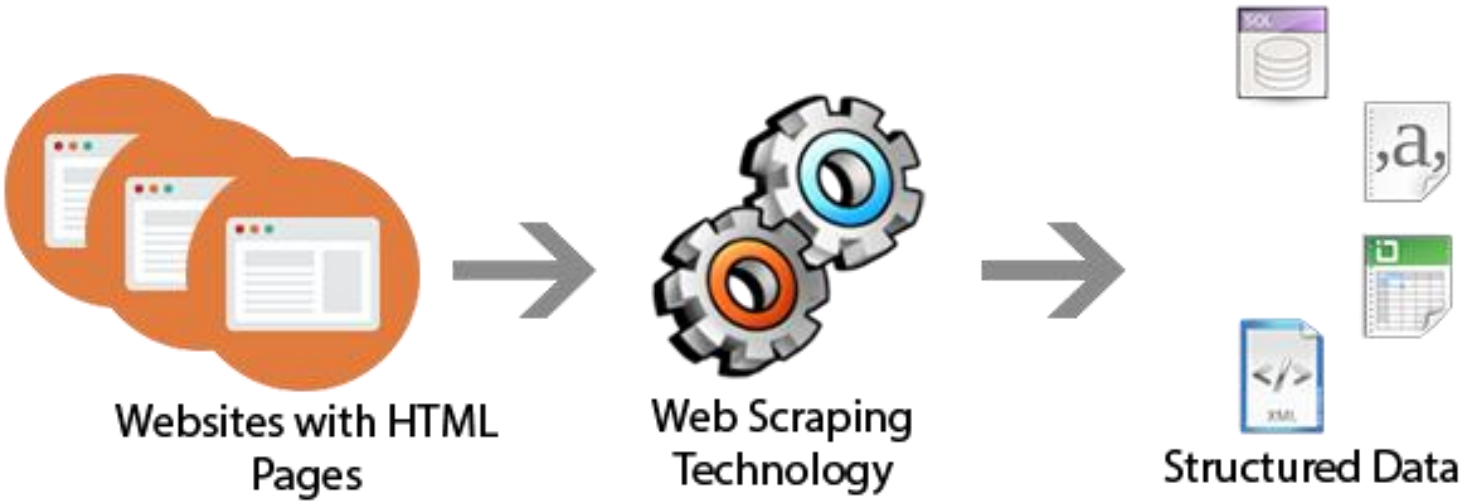
1BvBMSEYstWetqTFn5Au4m4GFg7xJaNVN2

Web crawling



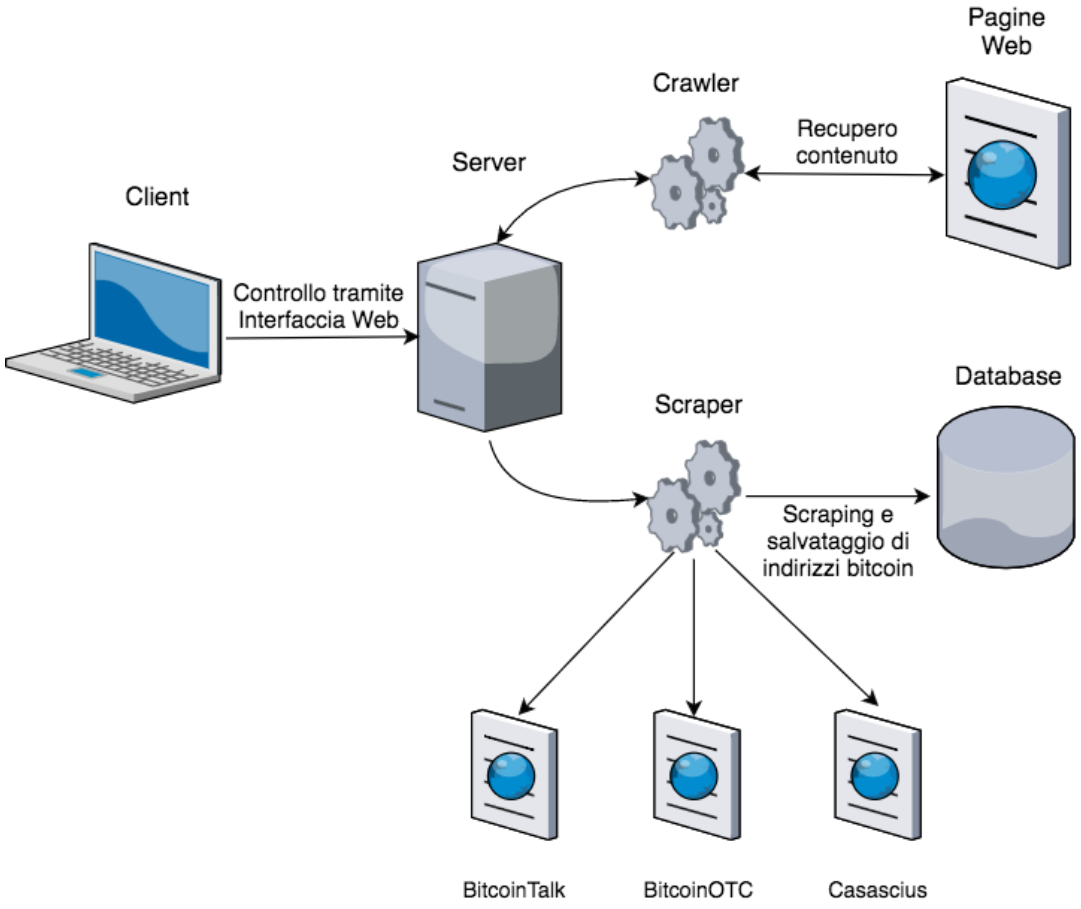
Il web crawling è il processo di analisi di siti web utilizzato per indicizzarne tutti i contenuti. Un crawler, anche conosciuto come spider, è un software specializzato per prelevare tutto il contenuto di una pagina web e seguirne i vari link per analizzare siti web collegati o pagine secondarie.

Web scraping



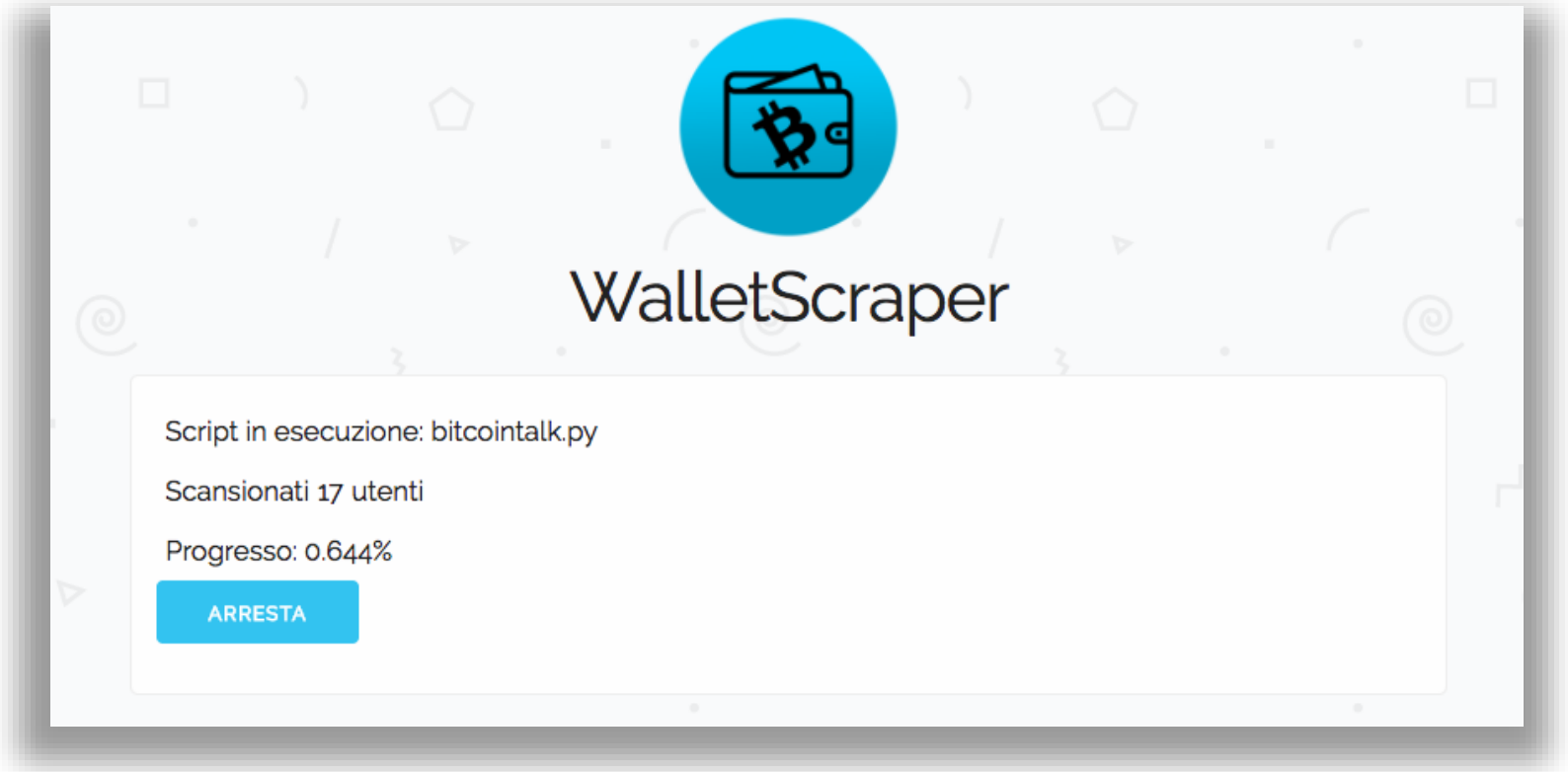
Il web scraping, chiamato anche in diversi modi tra cui *web data extraction*, è una tecnica, di solito automatizzata, che consiste nel prelevare singoli dati da un insieme di pagine web, per collezionarli all'interno di database o file per un'analisi futura.

Our crawling/scraping application



Esecuzione dello scraper

Esecuzione di script python



Scraping mirato su siti web “interessanti”

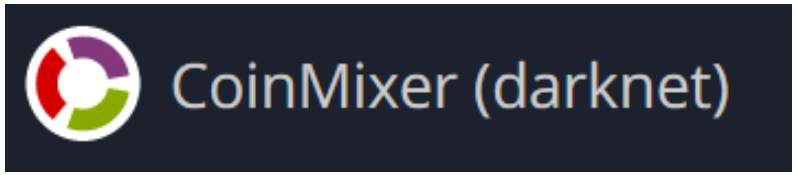
- BitcoinTalk
- Bitcoin-OTC
- Casascius

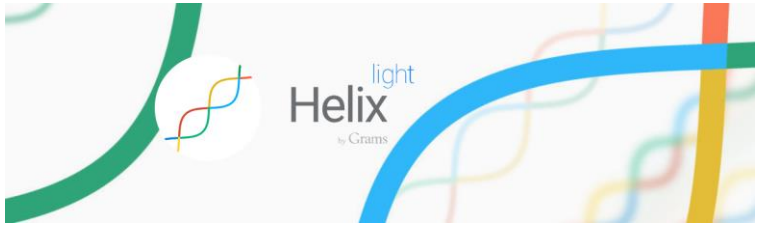
Un framework modulare: mixing service

Il nostro progetto

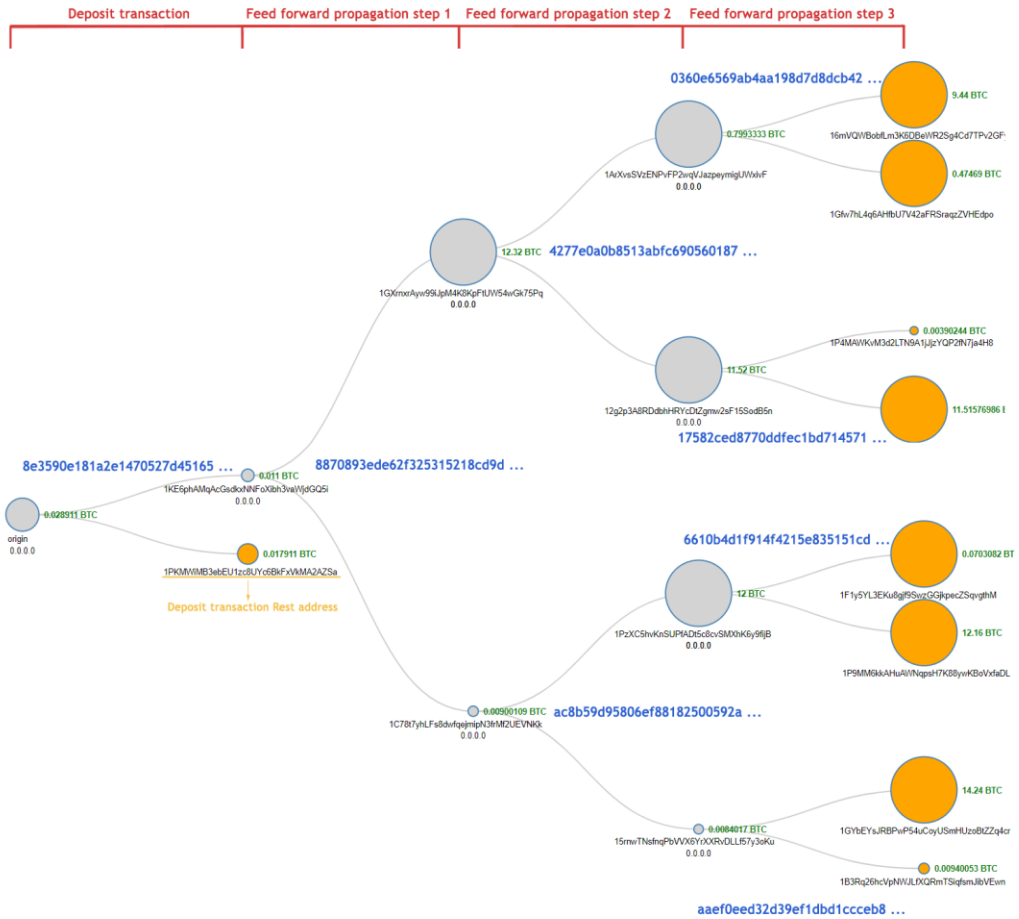
Trovare pattern comportamentali, indirizzi e transazioni appartenenti ai mixing services

partendo da transazioni campione effettuate con i seguenti mixing services:





Mixing service adresses



CoinMixer pattern

14 transazioni appartenenti al mixing service CoinMixer, accomunate dalle seguenti caratterisitiche:

- numero di output address nel range [2530-2534], altamente sovrapponibili
- recuperate tutte nel passo di estrazione K = -3 (Backward Propagation)
- effettuate in 14 giorni consecutivi (una al giorno)
- permettono il recupero di 7 transazioni "No Info" con le stesse caratterisitiche

transactions	TX1	TX2	TX3	TX4	TX5	TX6	TX7	TX8	TX9	TX10	TX11	TX12	TX13	TX14
TX1	100.0%	99.05%	98.62%	97.95%	97.12%	96.45%	96.09%	95.66%	95.18%	94.79%	94.43%	94.12%	93.76%	93.33%
TX2	99.05%	100.0%	99.17%	98.42%	97.59%	97.12%	96.68%	96.25%	95.78%	95.38%	95.03%	94.71%	94.23%	93.8%
TX3	98.62%	99.17%	100.0%	99.09%	98.15%	97.55%	97.12%	96.57%	96.09%	95.66%	95.3%	94.99%	94.47%	94.0%
TX4	97.95%	98.42%	99.09%	100.0%	98.78%	98.11%	97.63%	97.04%	96.49%	96.05%	95.7%	95.38%	94.83%	94.35%
TX5	97.12%	97.59%	98.15%	98.78%	100.0%	99.01%	98.46%	97.91%	97.32%	96.88%	96.45%	96.21%	95.62%	95.03%
TX6	96.45%	97.12%	97.55%	98.11%	99.01%	100.0%	99.25%	98.62%	98.07%	97.63%	97.2%	96.8%	96.33%	95.74%
TX7	96.09%	96.68%	97.12%	97.63%	98.46%	99.25%	100.0%	99.21%	98.7%	98.26%	97.83%	97.35%	96.76%	96.25%
TX8	95.66%	96.25%	96.57%	97.04%	97.91%	98.62%	99.21%	100.0%	99.13%	98.7%	98.26%	97.91%	97.16%	96.6%
TX9	95.18%	95.78%	96.09%	96.49%	97.32%	98.07%	98.7%	99.13%	100.0%	99.41%	98.97%	98.5%	97.83%	97.2%
TX10	94.79%	95.38%	95.66%	96.05%	96.88%	97.63%	98.26%	98.7%	99.41%	100.0%	99.45%	98.93%	98.14%	97.51%
TX11	94.43%	95.03%	95.3%	95.7%	96.45%	97.2%	97.83%	98.26%	98.97%	99.45%	100.0%	99.37%	98.46%	97.83%
TX12	94.12%	94.71%	94.99%	95.38%	96.21%	96.8%	97.35%	97.91%	98.5%	98.93%	99.37%	100.0%	98.89%	98.18%
TX13	93.76%	94.23%	94.47%	94.83%	95.62%	96.33%	96.76%	97.16%	97.83%	98.14%	98.46%	98.89%	100.0%	99.01%
TX14	93.33%	93.8%	94.0%	94.35%	95.03%	95.74%	96.25%	96.6%	97.2%	97.51%	97.83%	98.18%	99.01%	100.0%

Bitcoin Blender pattern 1

Le seguenti transazioni di tipo "Mixing Service":

- d67f4dac38ffc03c921f7936ea00718754a0e5923c25dbb0b023f7114e4c4ade
- b17bd6d623511af6dfa484090940e3ec47267dff1714a3abc29287cfc4c906bc

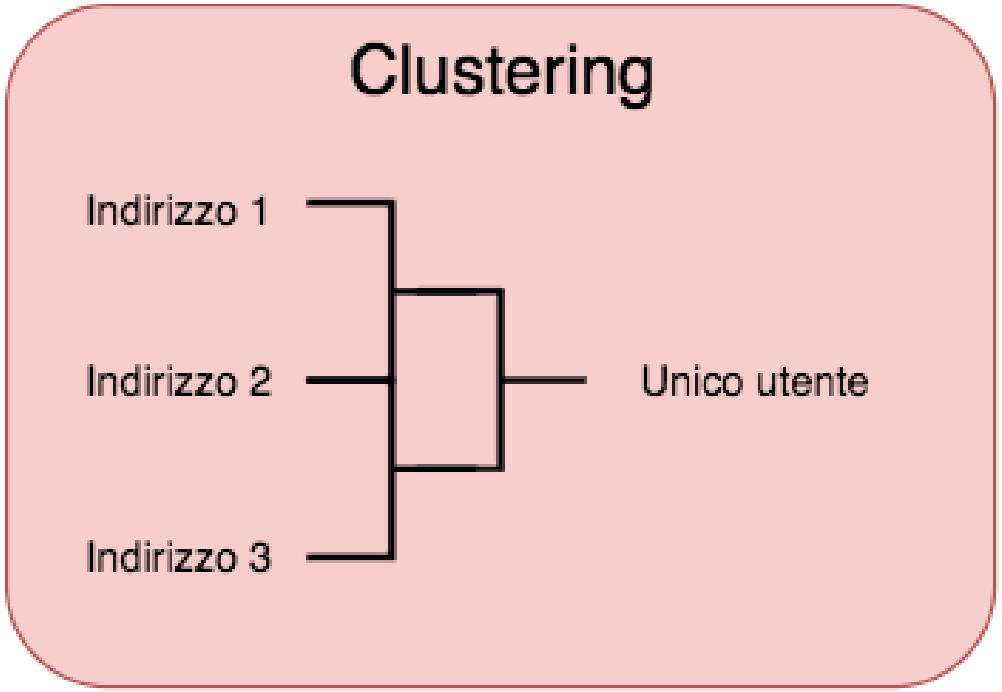
permettono il recupero di 2558 e 2560 transazioni "No Info", rispettivamente. Tali transazioni condividono lo stesso address sia fra gli input che fra gli output addresses (1N52wHoVR79PMDishab2XmRHsbekCdGquK).

Tale address viene utilizzato come accumulatore di bitcoin nelle transazione al passo K = 3 e come resto nelle transazioni al passo K = -3

Sommario	
Indirizzo	1N52wHoVR79PMDishab2XmRHsbekCdGquK
Hash 160	e71debe251bb26c7e757d9ae265da6e5d00f31b9
Utensili	Tag correlati - Uscite non utilizzate

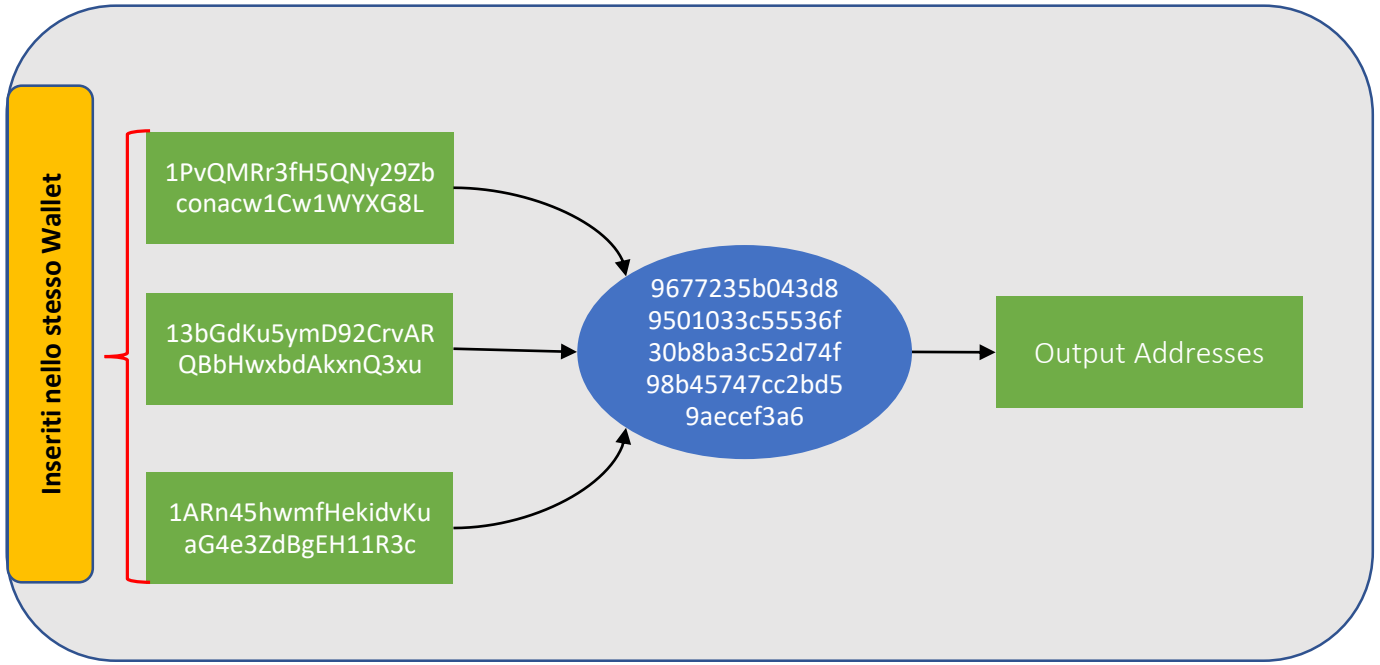
Le transazioni	
Nr. Transazioni	72619
Totale Ricevuto	1,484,636.91882993 BTC
Bilancio finale	5,440.88632347 BTC

Un framework modulare: clustering



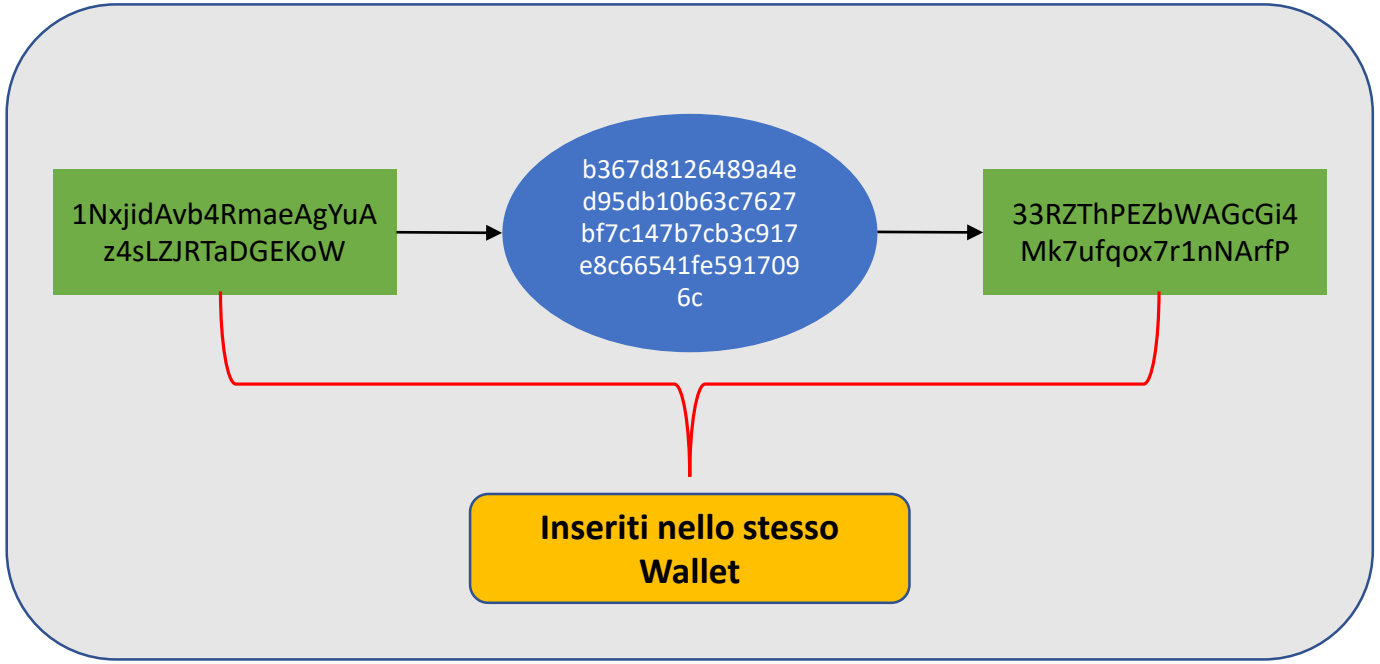
Multi-input heuristic

- La Multi-Input heuristic sfrutta le transazioni che presentano più di un indirizzo in input
- Questa euristica raggruppa gli indirizzi che compaiono in input nella stessa transazione
- Una volta applicato questo a tutte le transazioni della blockchain che soddisfano l'ipotesi, se due o più Wallet hanno degli indirizzi in comune vengono raggruppati formando un unico Wallet



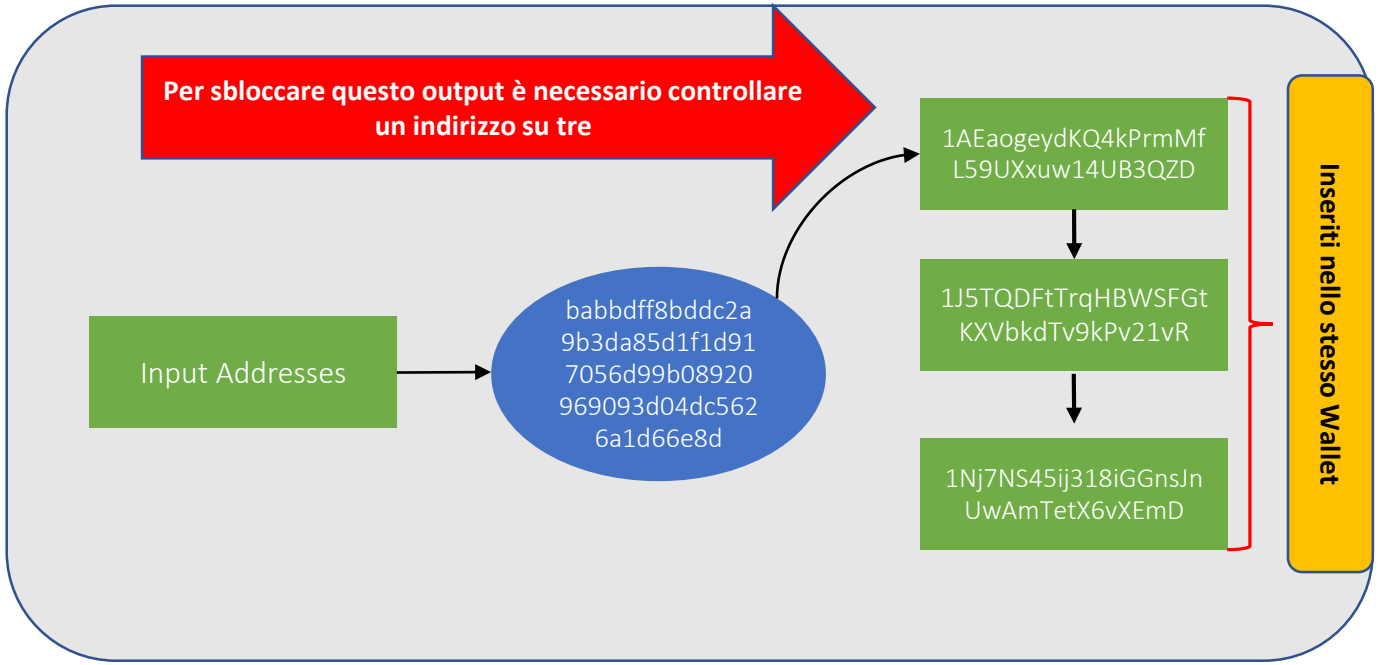
One-to-one heuristic

- Data una transazione avente un solo indirizzo in input e un solo indirizzo in output
- Dato un insieme di indirizzi **E**, contenente gli indirizzi appartenenti ad Exchange Services
- Se l'indirizzo di input e quello di output non compaiono all'interno dell'insieme E, allora appartengono allo stesso utente
- Una volta applicato questo a tutte le transazioni della blockchain che soddisfano l'ipotesi, se due o più Wallet hanno degli indirizzi in comune vengono raggruppati formando un unico Wallet



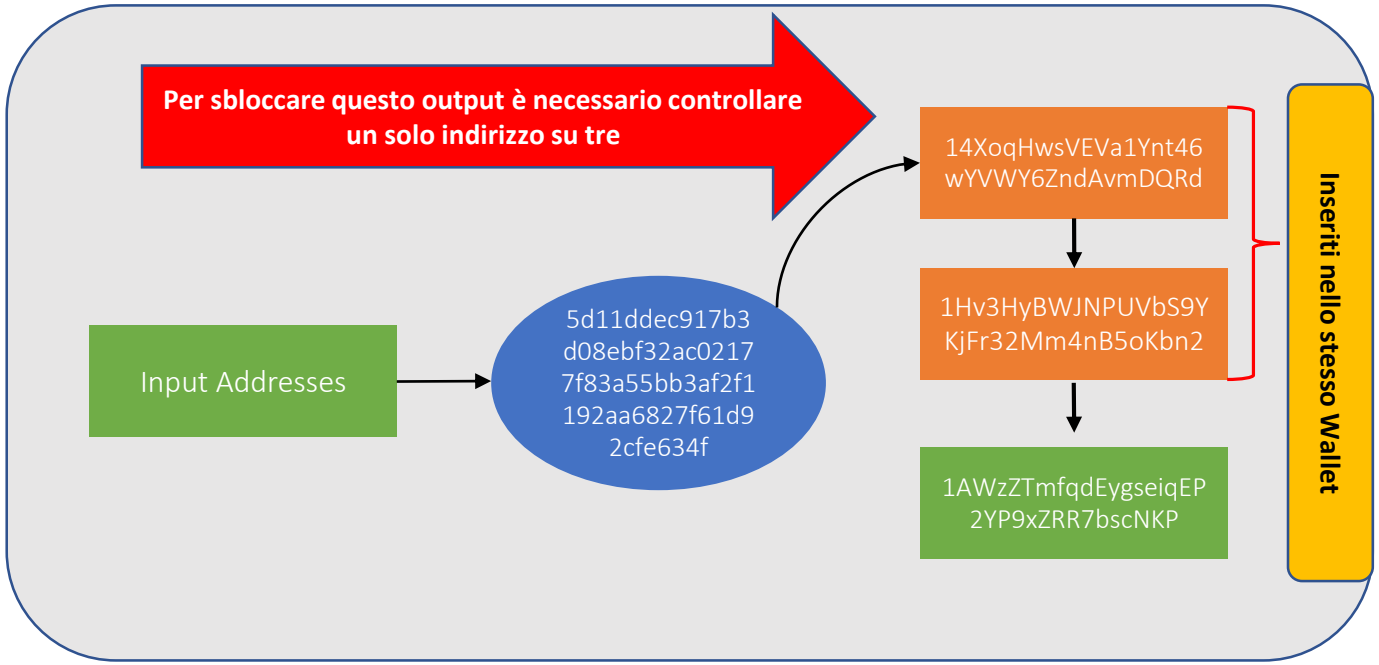
Multisig-one heuristic

- Questa euristica sfrutta gli output multisig 1-N o N-N
- In questo caso è un output 1-3, ciò significa che basta possedere la chiavi private su tre per poter riscattare l'output
- Per questo motivo applicando l'euristica a questa transazione i tre indirizzi dell'output multisig vengono inseriti all'interno dello stesso Wallet
- I tre indirizzi avrebbero fatto parte dello stesso wallet anche se fosse stato un output 3-3, in quanto per poterlo sbloccare si devono possedere le chiavi di tutti e tre gli indirizzi quindi verosimilmente appartengono allo stesso proprietario
- Come prima, una volta applicato questo a tutte le transazioni della blockchain che soddisfano l'ipotesi, se due o più Wallet hanno degli indirizzi in comune vengono raggruppati formando un unico Wallet



Multisig-two heuristic

- Questa euristica sfrutta gli output **riscattati** multisig M-N con 1>N e M<N
- In questo caso è un output 2-3, ciò significa che si devono conoscere due chiave privata di uno dei tre indirizzi per poter riscattare l'output
- Per questo motivo applicando l'euristica a questa transazione i due indirizzi che hanno sbloccato l'output multisig vengono inseriti all'interno dello stesso Wallet
- Ancora una volta applicato questo a tutte le transazioni della blockchain che soddisfano l'ipotesi, se due o più Wallet hanno degli indirizzi in comune vengono raggruppati formando un unico Wallet



Numero di indirizzi clusterizzati da ogni euristica

- Il numero totale di indirizzi presenti nel database è 115493058
- Quando si effettua una composizione di euristiche il risultato non è la somma degli indirizzi clusterizzati, poiché lo stesso indirizzo può essere clusterizzato in ognuna di esse
- Come si può notare, componendo tutte le euristiche di clustering si riesce a clusterizzare più del 76% degli indirizzi

Euristica	Indirizzi Clusterizzati	Percentuale di indirizzi clusterizzati rispetto al totale
MI	83867895	72,61%
O	5004254	4,33%
MS1	520396	0,45%
MS2	2263	0,001%
MI+O	87613567	75,86%
MI+MS1	84372511	73,05%
MI+MS2	83868035	72,61%
O+MS1	5523007	4,78%
O+MS2	5006384	4,33%
MS1+MS2	521263	0,45%
MI+O+MS1	88116265	76,29%
MI+O+MS2	87613699	75,86%
MI+MS1+MS2	84373211	72,61%
O+MS1+MS2	5523859	4,78%
MI+O+MS1+MS2	88116388	76,29%

Numero di Wallet trovati

- La cosa che salta subito all'occhio osservando i dati delle euristiche base è che la MS1 ha una media di indirizzi per wallet insolitamente grande rispetto alle altre
- Tutto ciò può stare a significare che, almeno fino al 2016, gli utenti che facevano uso delle transazioni multisig non erano molti e creavano ogni volta molti indirizzi per coprire il proprio indirizzo reale

Euristica	Numero di Wallet	Indirizzi per Wallet
MI	15839688	5,29
O	1348653	3,71
MS1	10900	47,74
MS2	918	2,46
MI+O	15046664	5,82
MI+MS1	15846740	5,32
MI+MS2	15839742	5,29
O+MS1	1357973	4,06
O+MS2	1349543	3,70
MS1+MS2	11229	46,42
MI+O+MS1	15052128	5,85
MI+O+MS2	15046713	5,82
MI+MS1+MS2	15847324	5,32
O+MS1+MS2	1358290	4,06
MI+O+MS1+MS2	15052173	5,85

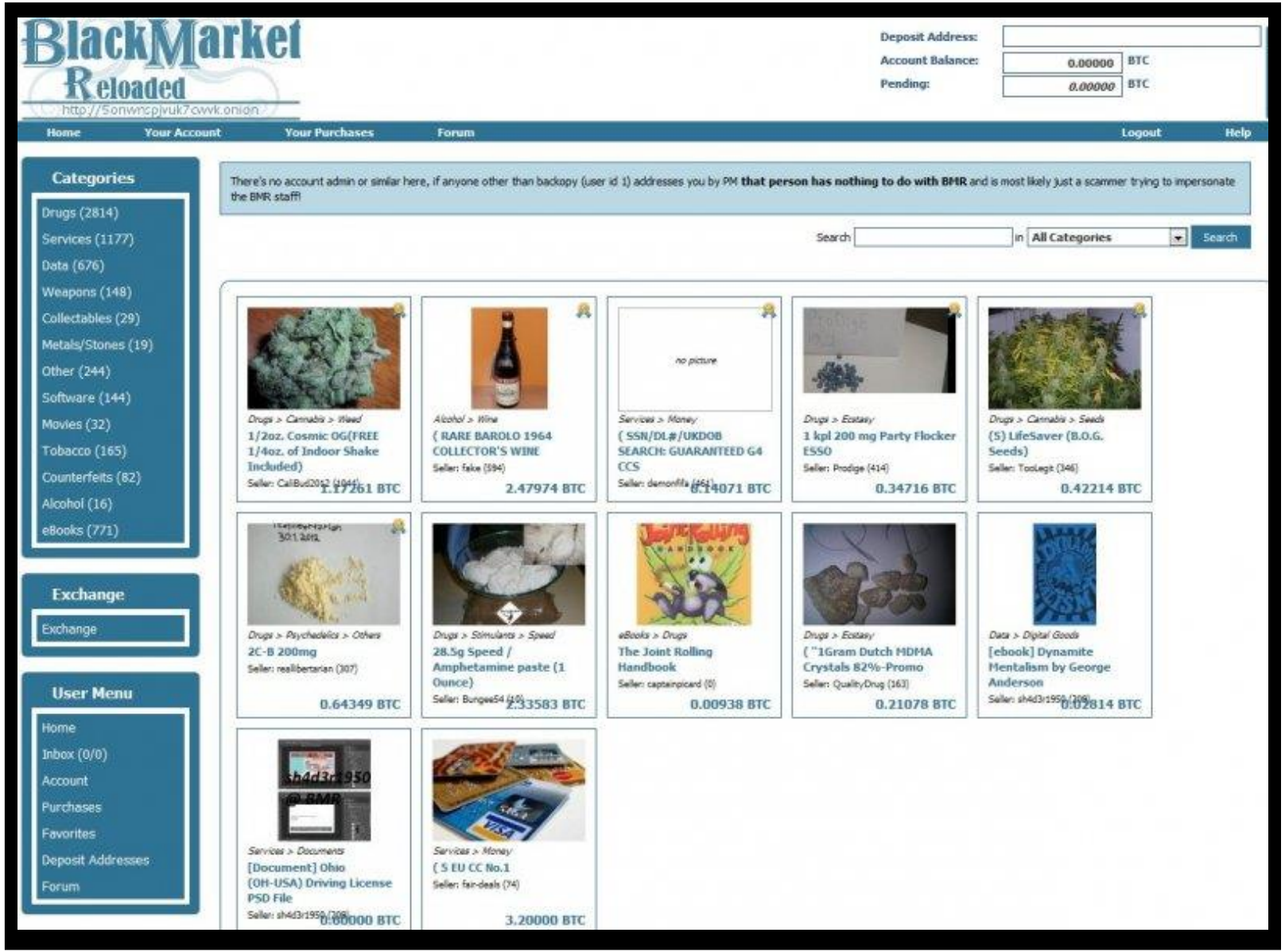
BlockChainVis

We propose ***BlockChainVis***, a tool dedicated to the visual analysis of Bitcoin transactions

Since the block-chain is an example of Big Data (more than 130 Gbyte raw), a straightforward visualisation in its entirety is not very significant (nor possible)

Hence, we have exploited some techniques from ***Visual Analytics*** (*VA*) to filter out undesired information, with the purpose to obtain a forensic-tool to efficiently and visually analyse the block-chain and help its investigation

BTC e BlackMarket



Progetto europeo
SEC-12-FCT-2016-2017

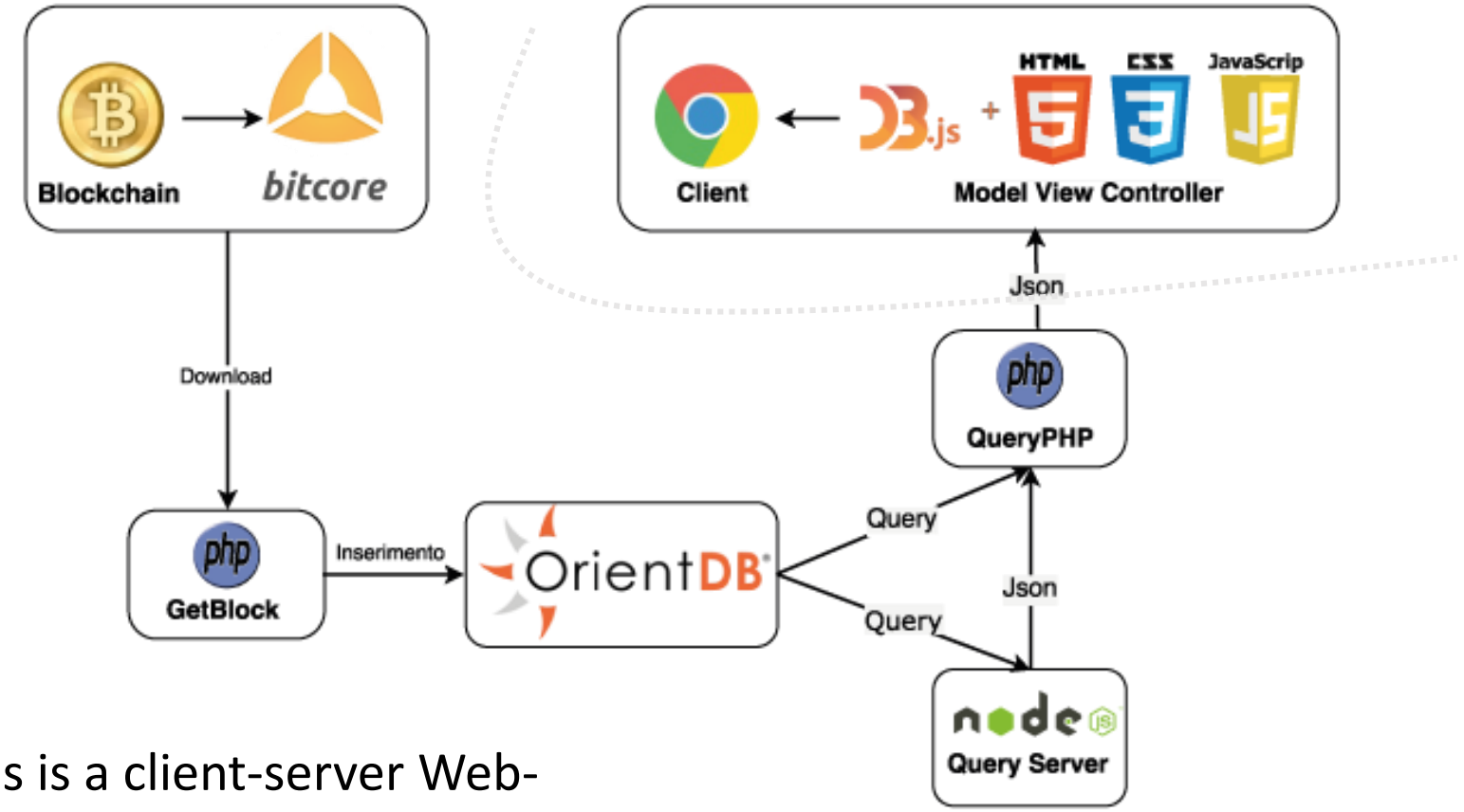
Specific Challenge:

Organized crime and terrorist organizations are often at the forefront of technological innovation in planning, executing and concealing their criminal activities and the revenues stemming from them. Law Enforcement Agencies (LEAs) are often lagging behind when tackling criminal activities supported by "advanced" technologies.

Sub-topic:

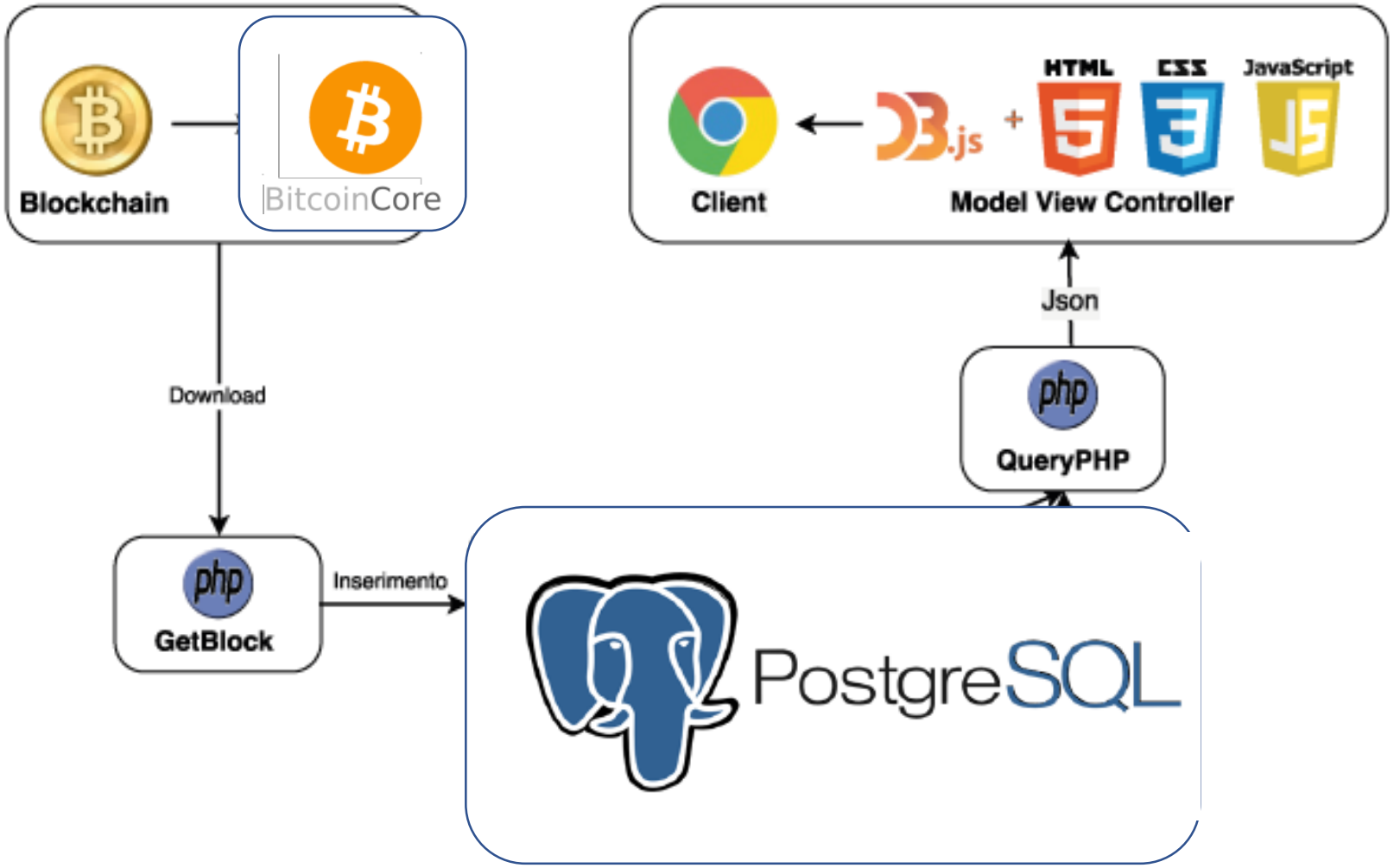
1.cyber-crime: virtual/crypto currencies des-anonymisation/tracing/impairing where they support underground markets in the darknet.

Architettura

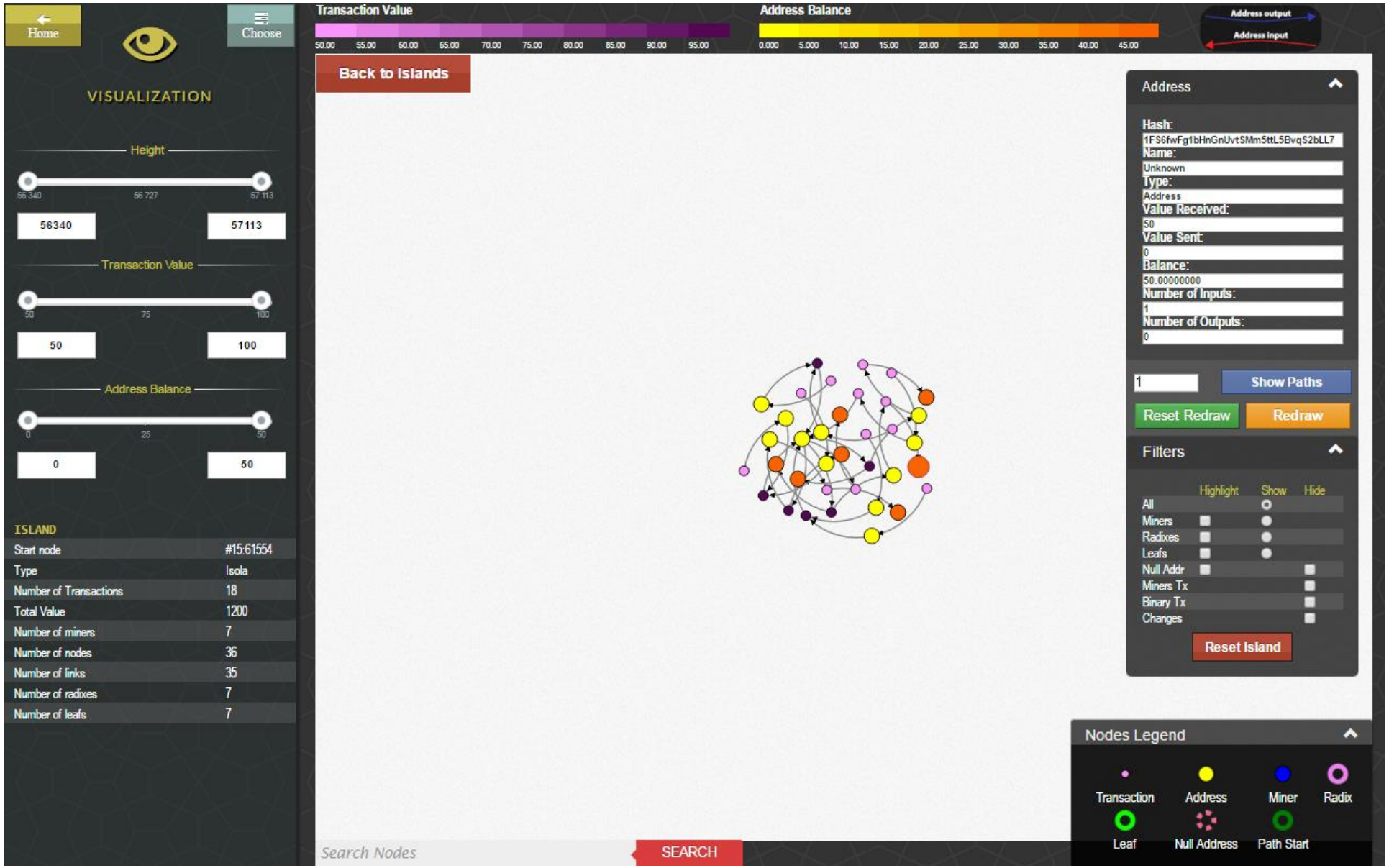


BlockChainVis is a client-server Web-application. It consists of a back-end (server-side) and a front-end (client-side).

Architettura II

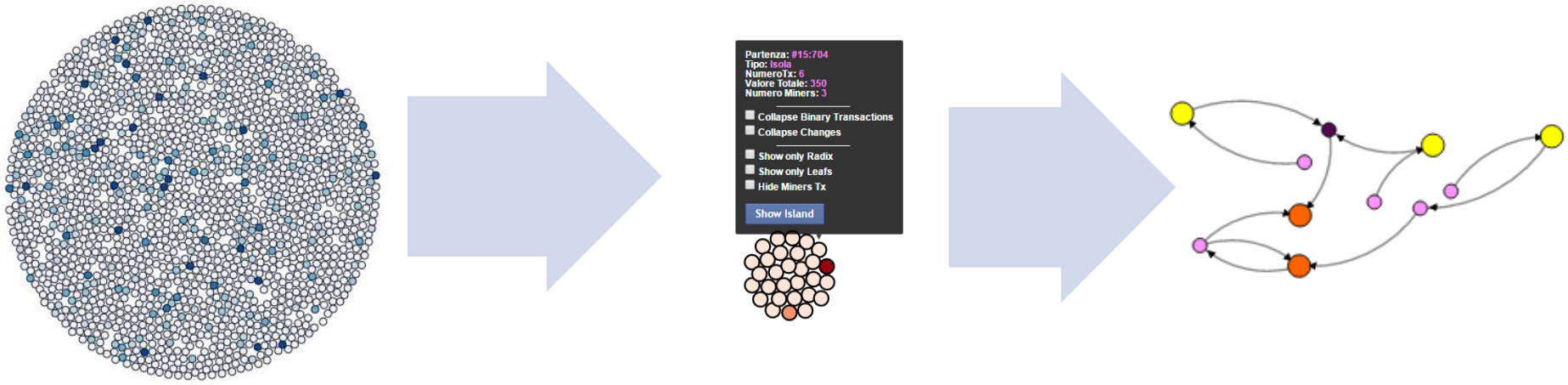



Home
Choose
VISUALIZATION
Choose a type
Please Choose...
Please Choose...
Single Transaction
Address Transactions
Archipelago
BLOCKCHAIN
Sezione di Visualizzazione



Interfaccia Applicativo con relativi strumenti

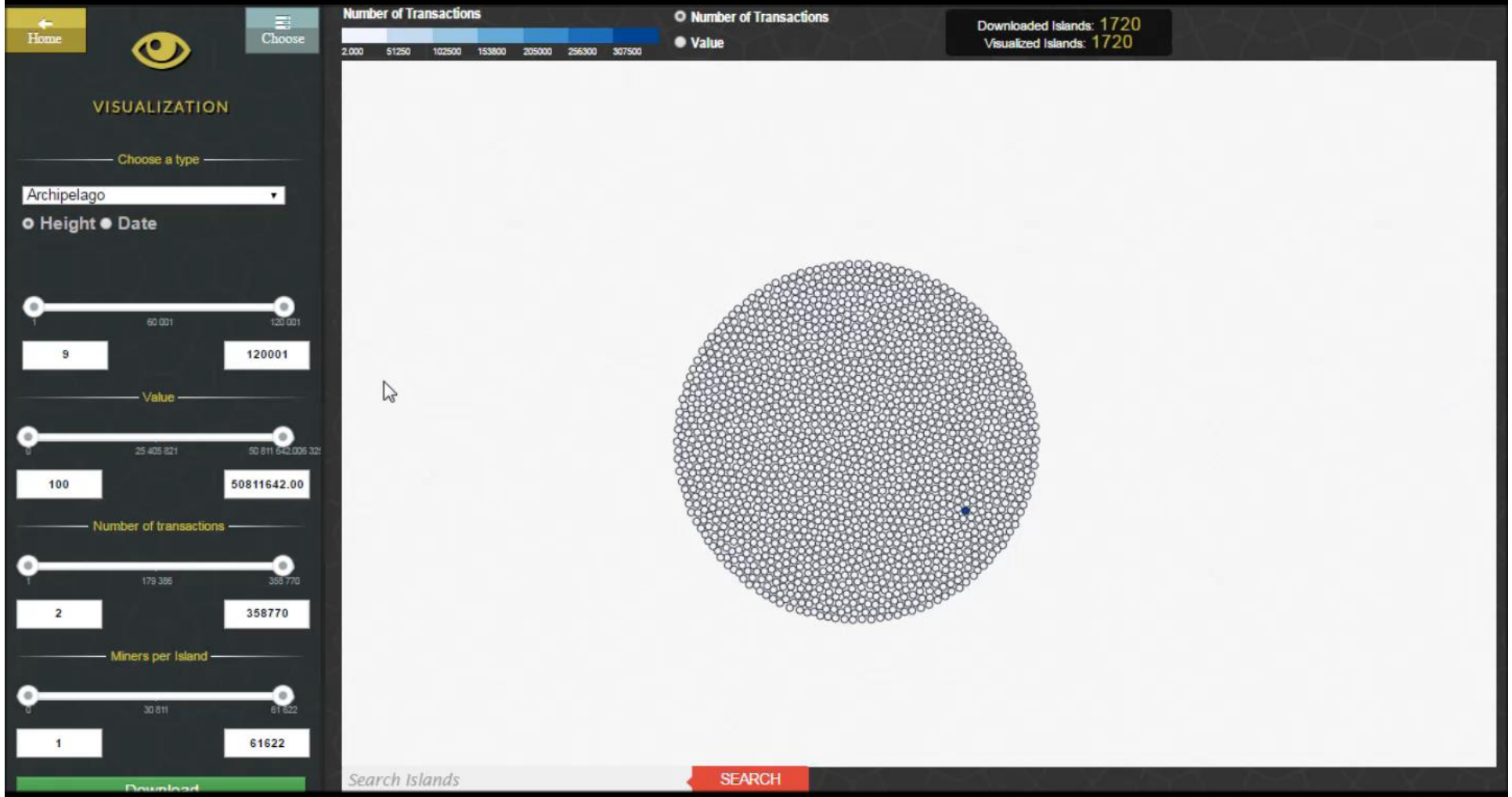
Composizione Layer di Visualizzazione



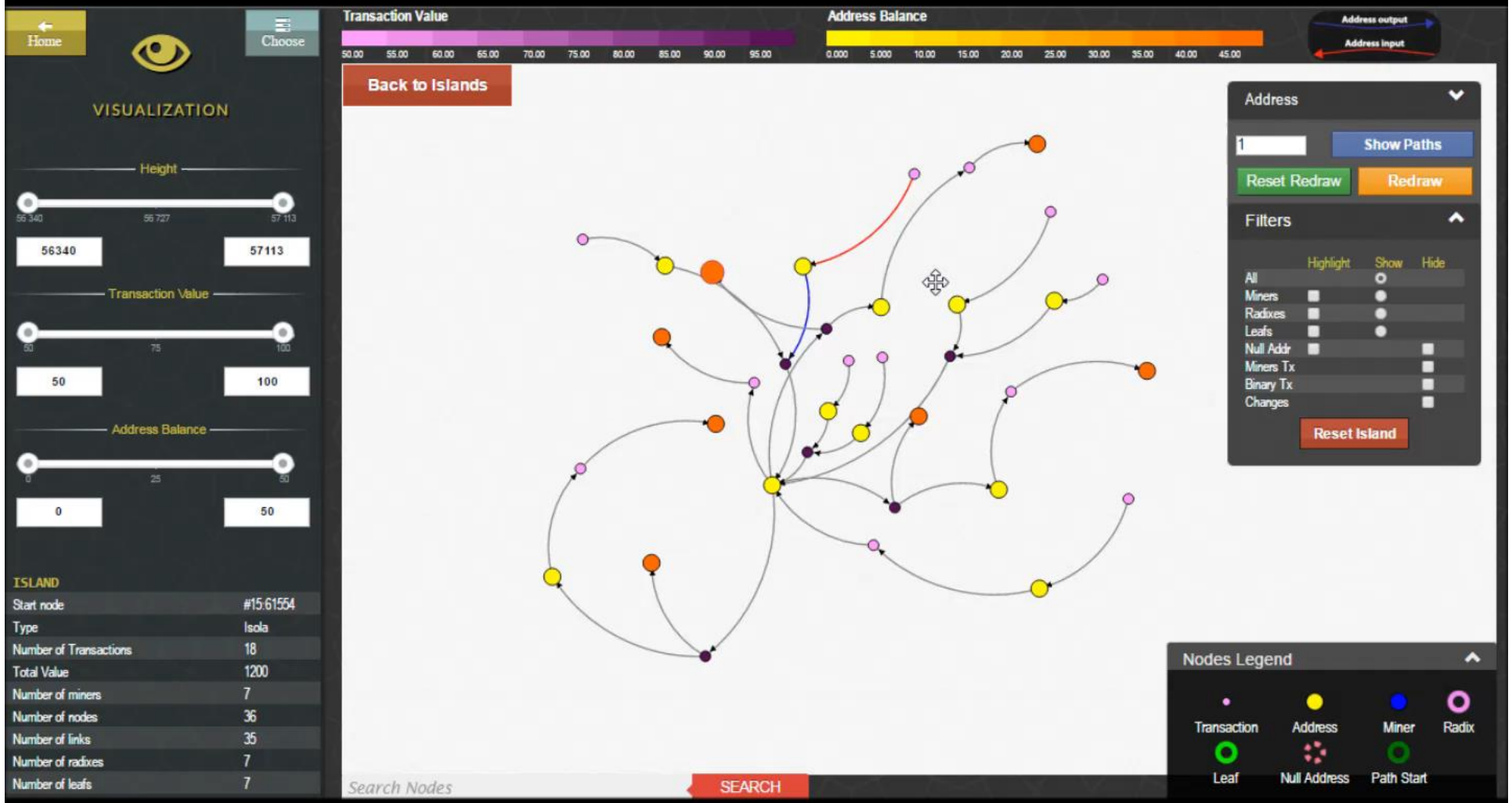
Arcipelago

Ulteriore raffinamento

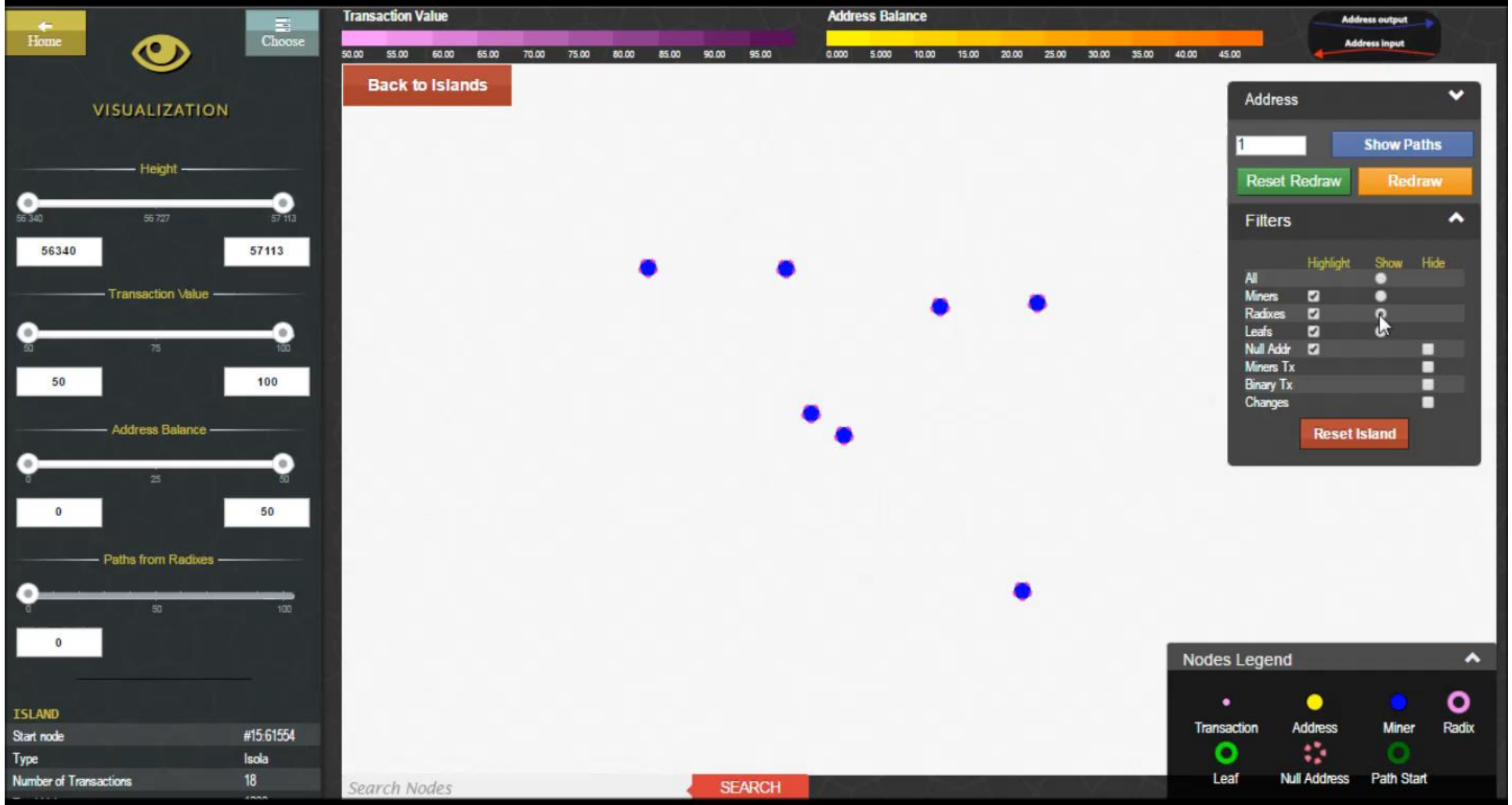
Isola



Layer dell’Arcipelago con relativi filtri

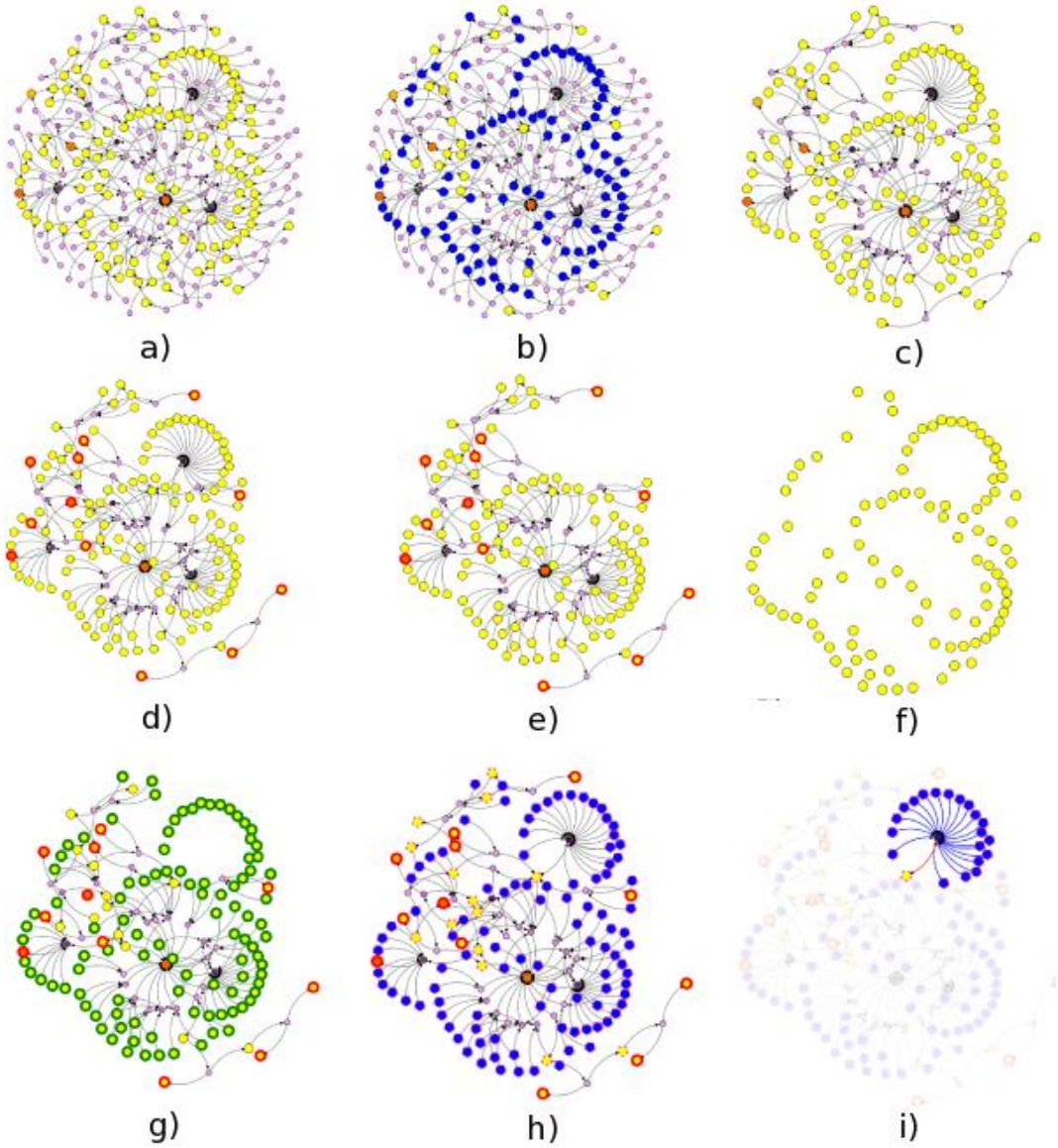


Evidenzia Miners



Visualizza solamente ‘’Radici’’

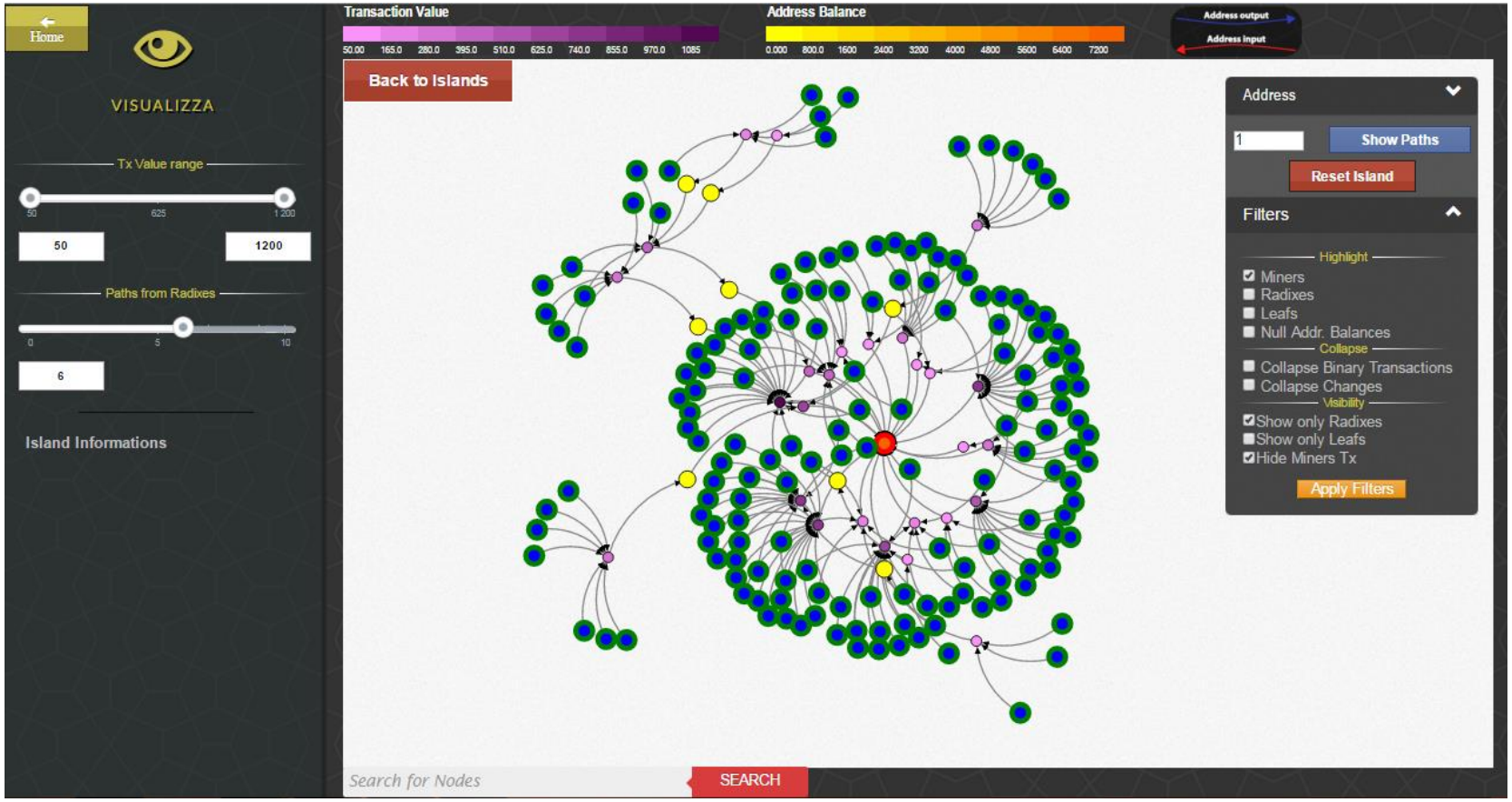
Island filters



An example of filter application on

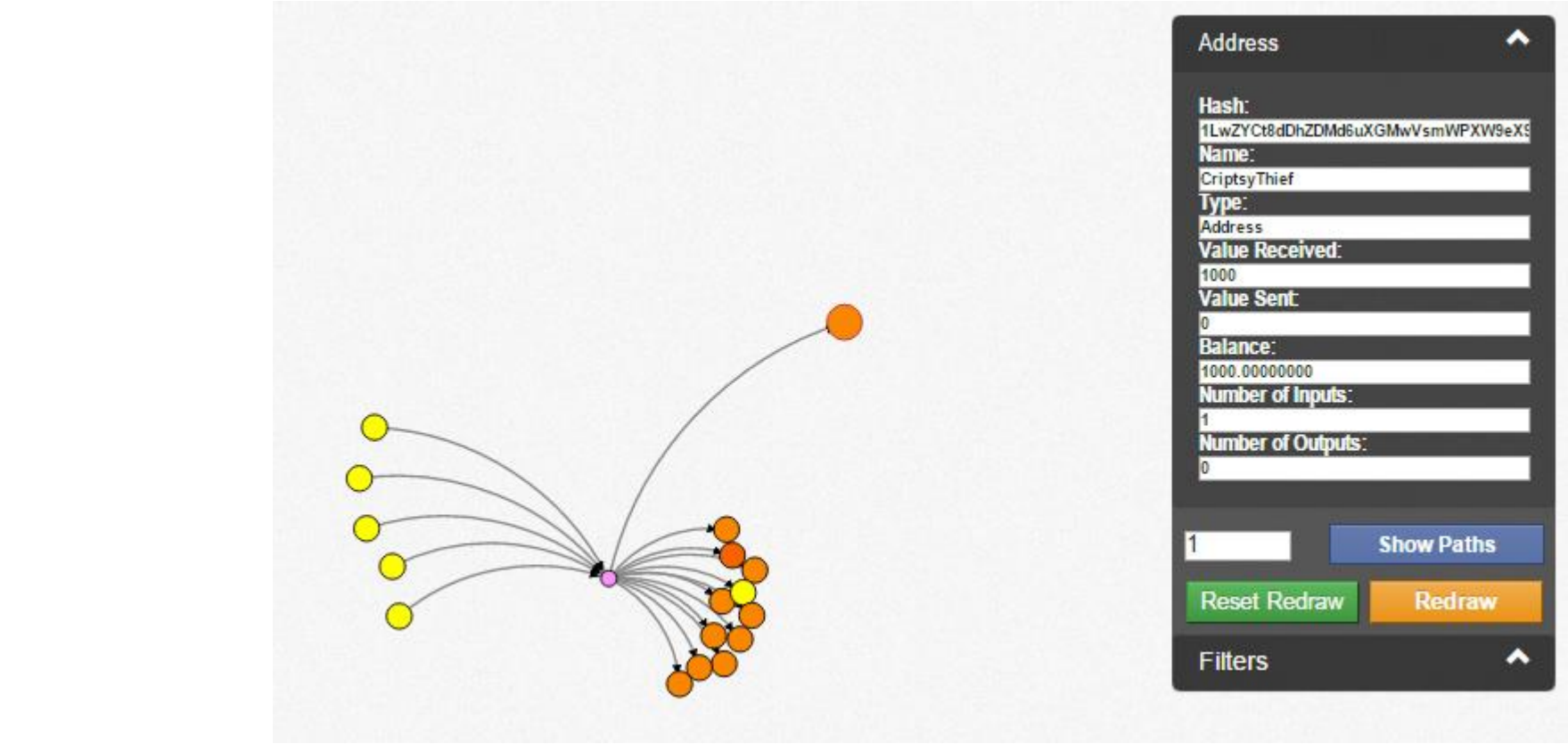
(a) the initial graph:
(b) by highlighting miners,
(c) hiding coinbase transactions (transactions rewarding the miners),
(d) highilighting leaves,
(e) a transaction-value interval,
(f) showing only roots,
(g) visualising paths (most of the green/yellow nodes are roots and the others are leafs),
(h) combines *b* and *f* together (darker nodes highlight the same miners in (b)),
(i) focuses on a given transaction.

Pool of Miners



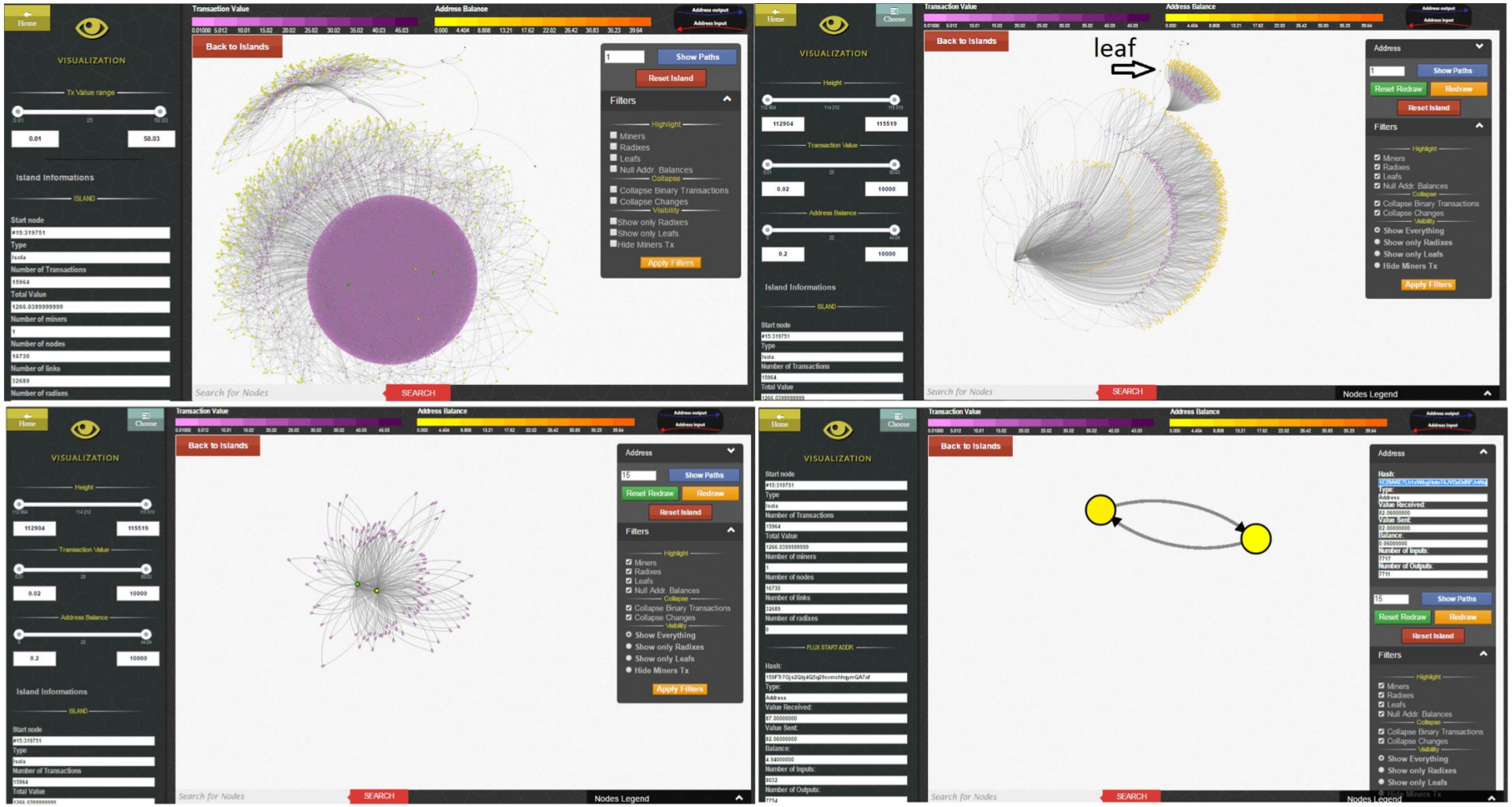
Cryptsy.com

Hacked in July 2017, 13,000 bitcoins stolen

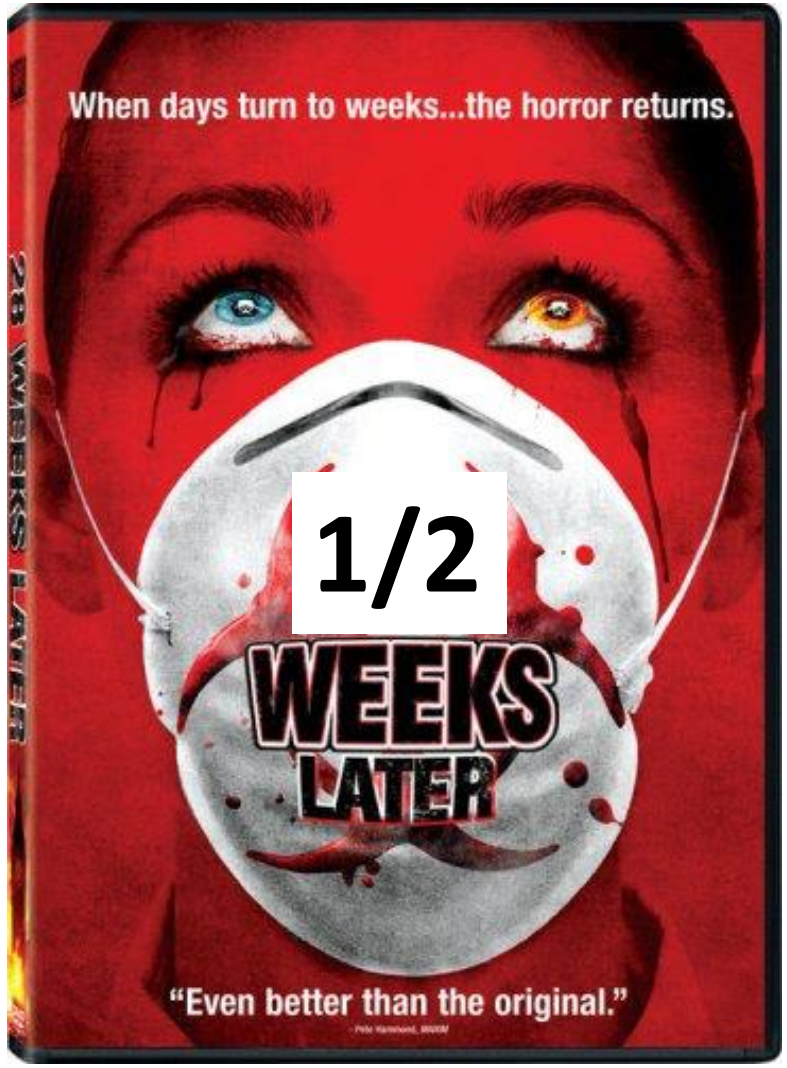


http://www.coindesk.com/cryptsy-ceo-millions-digital-currency-steal/

One-to-one exchange



WannaCry one week later



WannaCry one week later

WannaCry (alternatively WannaCrypt, WanaCrypt0r 2.0, Wanna Decryptor) is a ransomware computer-worm that targets the family of Microsoft Windows operating systems.

It was first discovered on May 12th 2017 (Friday)

A 22-year-old web security researcher from North Devon in England known as MalwareTech discovered kill switch by registering a domain name he found in the code: iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com.

Within a day was reported to have infected more than 230,000 computers in over 150 countries

WannaCry

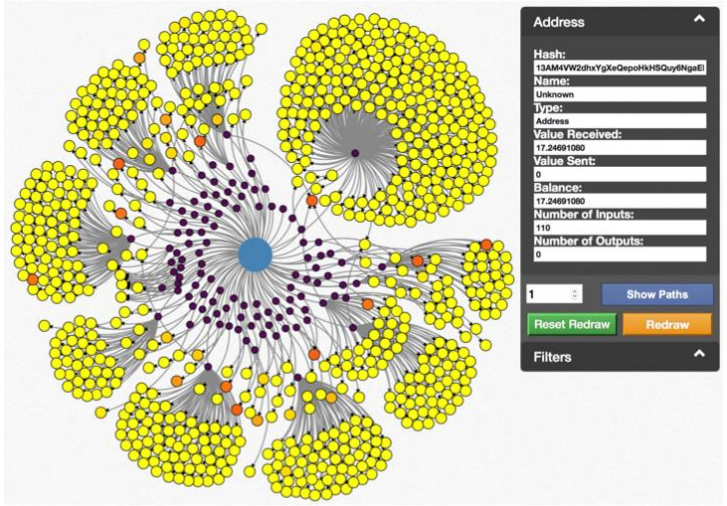
WannaCry propagates using EternalBlue, an exploit of Windows' Server Message Block (SMB) protocol

It demands a ransom of 300$ in bitcoins at the time of infection, which doubles to 600$ after three days.

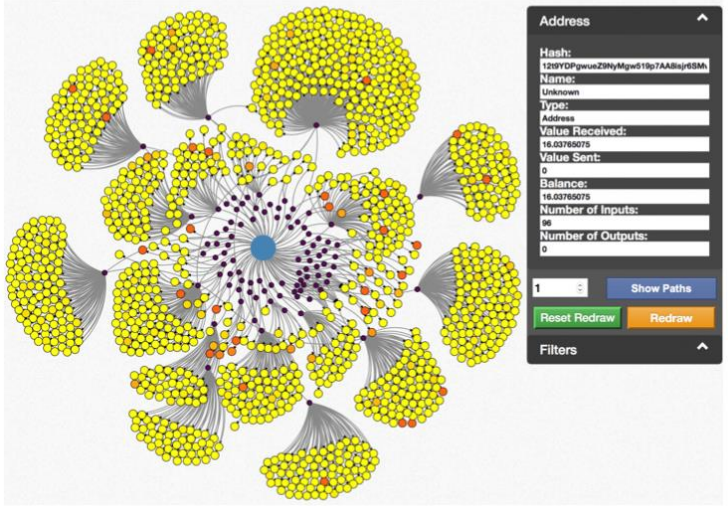
In BlockchainVis, we consider all the transactions mined between 2017-05-12 00:21:05 and 2017-05-18 09:42:34

We filter out all the inputs of the transactions, in order to only focus on final money destinations

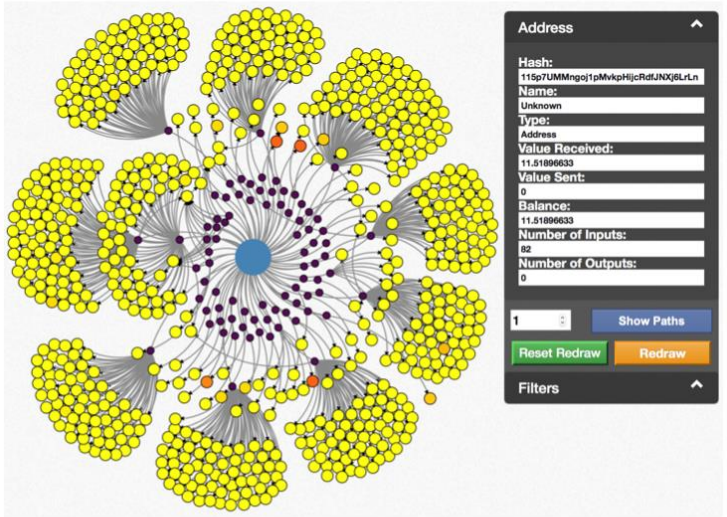
WannaCry



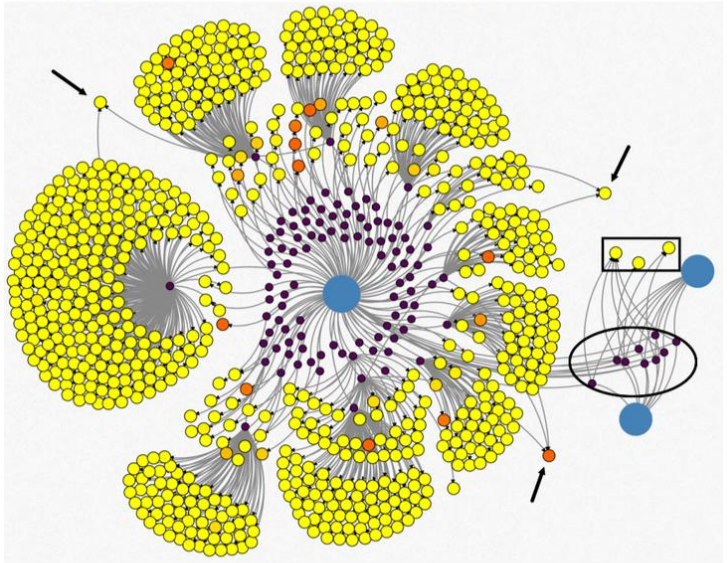
(a) 13AM4VW2dhxYgXeQepoHkHSQuy6NgaEb94.



(b) 12t9YDPgwueZ9NyMgw519p7AA8isjr6SMw



(c) 115p7UMMngoj1pMvkpHijcRdfJNXj6LrLn.



(d) Highlighting the three addresses on Fig. 1a.

Considerations

The sum of money moved to such three addresses during the first week is 17.247 + 16.037 + 11.518 = 44.802B.

The total number of inputs that concerns all three of them is 110 + 96 + 82 = 288

They transfer an amount of bitcoins very close to 300$ or 600$: usually between 0.15B and 0.18B (resp., 0.3B-0.36B)

One paid a higher ransom: 1.99B, which correspond to 3, 372$ (11 infected machines)

Considerations

No payment before 12th of May, only 45 payments more during the second week (89% of money collected first week)

No outbound transaction from all the three addresses

85 of such 333 transactions are less than 0.01B (~ 2$).

Victims who paid ransom after 2 weeks are 333-85 = 248

Flowers of payments

“Flowers with many petals”; for instance, there are 12 of such many-output transactions for the first address (largest has 246 outputs)

Most of these outputs receive a low amount of bitcoins, while a few addresses receive more than the ransom. The largest of them moves a total amount of 147.83B, but 144 addresses (out of 246) receive less than 0.1B.

Six of these receivers belong to Poloniex.com, a US-based cryptocurrency exchange and lending service provider. Some other addresses refer to different betting, investing, or wallet services, e.g. Cubits.com

Hidden messages

No ransom was split among multiple addresses

But a set of 9 transactions (grouped by an ellipses) moved some bitcoins to all the three WannaCry addresses. The amount is always less than 1$

Messages sent to WannaCry addresses to reach high visibility

Five of them only have one input address, but all 5 addresses contain the string “1DoDiK” (president of Rep. Srpska?)

One has four input addresses, each of them containing part of:
“You are a c**t”

Future Work

- Automatic filters (a combination of filters)
- Temporal Analysis
- Continuous Analysis
- Automatic notification

We are improving the hardware

- Azure for Research Award “Data Science for investigating the block-chain in Bitcoin”

- Amazon AWS Research Grant

A high performance server

Thanks for your attention

Contacts:

stefano.bistarelli@unipg.it

francesco.santini@unipg.it