

# Unveiling the Black Box

Merkle Science Attribution Methodologies

Last Updated: 10 October, 2023



# Table Of Contents

## Definitions

## Problem Statement

## Current state of Database

## Attribution Methodology

### 1. Dusting

- a. Entity Dusting Procedure
- b. Dusting Procedure for Nested Exchanges

### 2. Clustering Mechanisms

- a. Clustering for UTXO-based blockchains
- b. Clustering for EVM-centric blockchains

### 3. Proof of Reserves

### 4. Blockchain Explorers

### 5. Partnerships and Law Enforcement Collaborations

### 6. OSINT and In-house Investigations

### 7. Data Validation Procedures

## Definitions

**Blockchain Explorer** - A software for visualizing blocks, transactions, and blockchain network metrics (e.g., average transaction fees, hashrates, block size, block difficulty).

**Blockchain Forensics** - Blockchain forensics is the study of investigating criminal activities on the blockchain to trace, analyze, and identify illicit transactions, suspicious behaviors, and individuals involved in criminal activities.

**Chain Peeling** - Chain peeling a technique used to launder large amounts of illegally-obtained cryptocurrency by initiating a long series of small transactions.

**Clustering** - Clustering is a technique used to de-anonymize blockchain data, connecting multiple wallets associated with a common user or entity.

**KYC** - KYC, short for 'know your customer,' is the process through which financial institutions are mandated to conduct specific identity and background verifications on their customers before granting access to their products or platforms. This procedure is an integral component of the comprehensive regulatory measures adopted globally to combat money laundering.

**UTXO** - An unspent transaction output (UTXO) signifies a transaction output that remains available for use as input in a subsequent transaction. The UTXO model serves as a foundational component in Bitcoin and various other cryptocurrencies.

**Proof of Reserves** - Proof of Reserves (PoR) is an auditing method for crypto firms to demonstrate solvency by verifying customer assets match the company's reserves

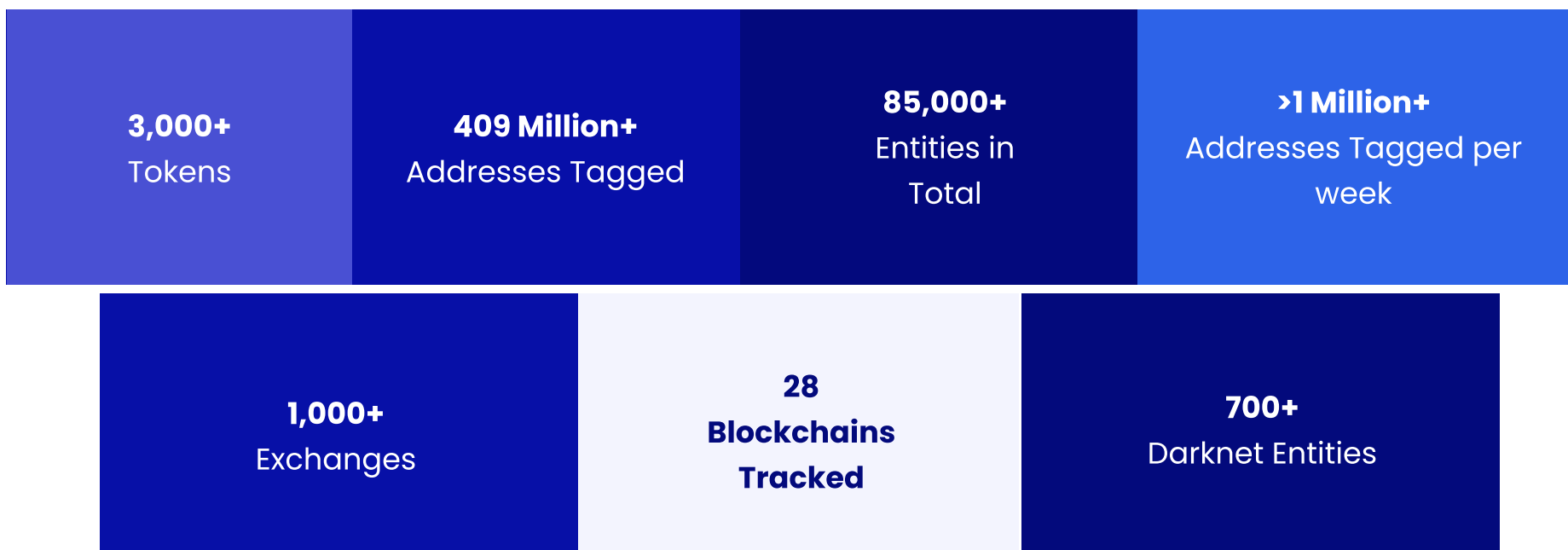


## Problem Statement

According to recent industry discourse, there's a pressing concern regarding the credibility of blockchain data as evidence in legal proceedings. The core of the debate revolves around the reliability and transparency of attribution mechanisms in blockchain forensics. The ramifications of this issue are profound, as it challenges the integrity of the data. Erroneous attributions can lead to tracing mistakes, amplifying the risks on wallet addresses leading to negative labeling, potentially resulting in incorrect arrests, and misguided legal decisions.

This report provides a concise insight into our data collection and attribution methods. With a focus on transparency, we aim to arm law enforcement with the necessary clarity for effective case-building using our data. Part 1 of the report presents the current status of our database. In Part 2, we spotlight our core attribution mechanisms, underscoring our dedication to accuracy and transparency.

## Current State of Database



## Complete Coverage

## Lite Coverage



This list encompasses both complete and lite coverage.

## PART 2. Advancing Transparency in Blockchain Attribution

Given the complexity of emerging obfuscation techniques, it's essential to validate blockchain-derived evidence with precise attribution. Upholding a commitment to transparency, we present a comprehensive yet incisive breakdown of our attribution and clustering techniques. In doing so, we seek to enhance law enforcement's ability to seize assets and secure convictions while making the process of parallel reconstruction of an investigation more efficient.

### Dusting

In the context of cryptocurrency, dusting refers to the practice of strategically sending tiny but negligible amounts of cryptocurrency to a specific address. Once the funds are sent to the address, by tracking the movement of these "dust" amounts, we determine associated clusters and the active hot wallets of exchanges.

Central to this technique is the role of the deposit address, a unique identifier generated by an exchange platform to facilitate the receipt of cryptocurrency. It directs where incoming funds should be sent. Upon receiving funds at this address, they are typically moved to the exchange's primary hot wallet — thus revealing other addresses belonging to the exchange as well.

The versatility of the dusting technique allows its application across multiple blockchain networks and applies to other service provider types as well.

Dusting is a common practice in blockchain analytics to determine if multiple addresses belong to the same entity, shedding light on possible dusting attacks or other malicious activities.

### Dusting Procedure

#### 1. Account Creation:

Create an account on the target exchange.

#### 2. KYC Compliance:

Complete any mandatory KYC procedures.

#### 3. Obtain a Deposit Address:

Request a deposit from the exchange, which, in turn, provides a deposit address.

#### 4. Send Cryptocurrency:

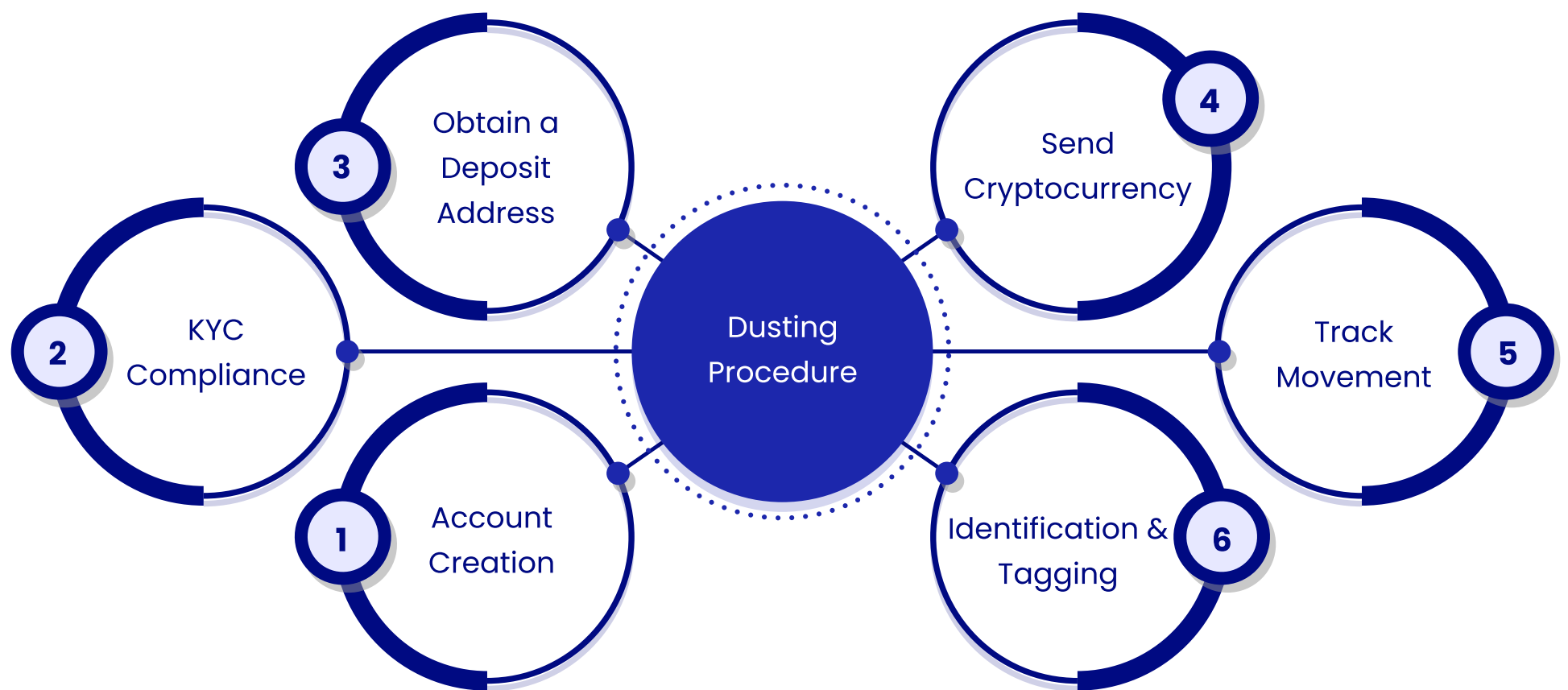
Dispatch small cryptocurrency amounts to the obtained deposit address.

#### 5. Track Movement:

Monitor the trajectory of these funds. Funds migrate to the platform's hot wallet.

#### 6. Identification & Tagging:

If funds move to an untagged hot wallet, that wallet's address is collated and tagged in our database



## The Entity Dusting Procedure

On the contrary, for the funds already deposited in the exchange wallet, we apply withdrawal methodology to get the hot wallet address of the exchange

1. **Withdrawal Analysis:** Request a withdrawal from the exchange, by providing a withdrawal address.
2. **Collect & Tag:** Collect the hot wallet address from where the funds have moved to the withdrawal address.

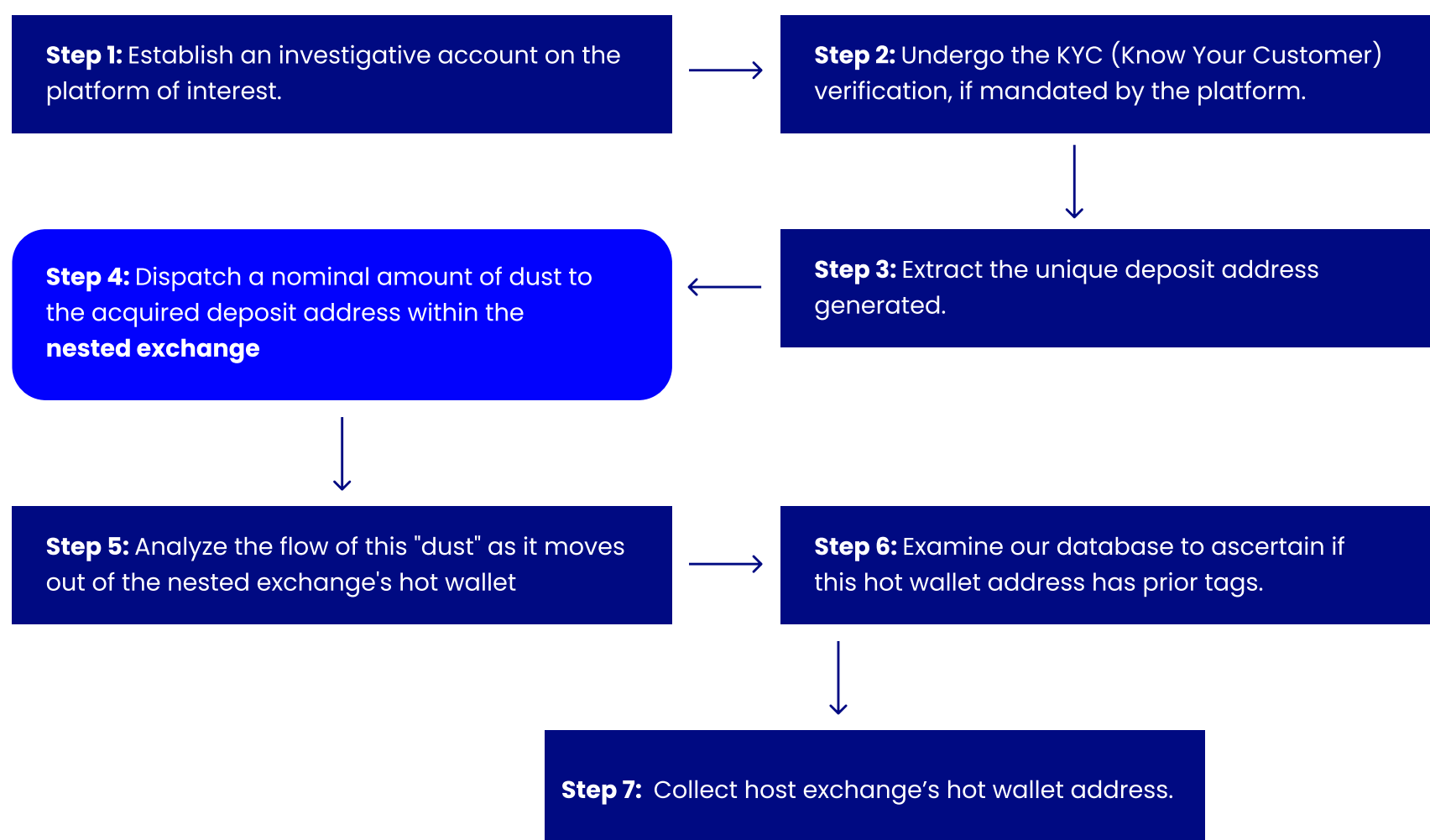
## Nested exchanges

The complexity increases with the involvement of nested services. In the context of cryptocurrency, nested services operate through intermediary platforms, allowing trades using accounts from another exchange. To address these complexities, we deploy investigative strategies and analyze fund movements.

**Note:** We undertake a validation process to ascertain whether the identified hot wallet belongs to any service provider by analyzing the operational workflow of the nested service.

# How Dusting of Nested Exchanges Work?

## How do we identify a nested service



## Accuracy and Oversight

The versatility of the dusting technique allows its application across multiple blockchain networks. By adjusting the method to the peculiarities of each network, we can achieve a comprehensive view of the hot wallets operating in different blockchain ecosystems.

The accuracy of this method is unparalleled for prominent exchanges, boasting a 100% accuracy rate. After the execution of each transaction, our intelligence team carefully monitors fund movements, which is why, this process gives an entity-level precision.

## Clustering

Blockchains provide information about transactions between addresses (similar to a bank account), including the amounts sent and received, and the timestamp. However, the entities, particularly – Virtual Asset Service Providers (VASPs) executing these transactions, remain unidentified. Merkle Science equips law enforcement agencies to identify the entities associated with an address or, more often, a group of addresses.

"Clusters" denote collections of addresses presumed to be governed by a singular entity. Merkle Science harnesses a proprietary algorithm that analyzes the transaction history to identify associated addresses – i.e. addresses that might be controlled by the same individual or entity.

The way we group or "cluster" these addresses depends on the type of blockchain:

- **UTXO Model**
- **EVM**

**UTXO Model:** Used by networks like Bitcoin and Litecoin, the Unspent Transaction Output (UTXO) model represents fund ownership by tracking transaction outputs that haven't been spent.

### Common Spend Heuristics

The common-input-ownership heuristic, also referred to as the co-spend heuristic, posits that all inputs within a transaction are attributed to a single entity. Multiple-input transactions arise when a user's wallet lacks a single Unspent Transaction Output (UTXO) of sufficient value to cover their payment, necessitating the use of multiple inputs. In cases where different addresses contribute multiple inputs to a single transaction, Merkle Science groups these sets of input addresses together to indicate common ownership by the same entity. Upon discovering a new address that co-spends with addresses already associated with an existing cluster, the new address is methodically integrated into the pre-existing cluster, thereby categorizing it under the same entity.

**Note:** This is one example of the various clustering heuristics we employ for blockchains based on the UTXO model.



UTXO-based Clustering Mechanism



## Summary

USD BTC

This transaction was first broadcast to the Bitcoin network on July 09, 2021 at 12:06 PM GMT+5:30. The transaction currently has 46 confirmations on the network. At the time of this transaction, 156.11445185 BTC was sent with a value of \$5,139,808.33. The current value of this transaction is now \$5,235,630.67. [Learn more about how transactions work.](#)

Hash	Amount	Address	Amount
ca7eb14be201a0718eb5cd4ddca3e0ad5816d5a59c954f5816a0...			2021-07-09 12:06
bc1qtg8gcfewcvj6js0qjfd8vxqu0wju67rmuyhmta	0.00899411 BTC		1PyYfLuLNAzAN3cKmVa7d4Uk17sihM8BaA 0.46875400 BTC
1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	51.66585879 BTC		34UqggprdGFnis7aDbH1VWAVC29MrptMC3 0.01196465 BTC
bc1qv0msw6pfsy2t1pazs85ijuzc3x97t99psfa8kg	0.10359576 BTC		35889QHjT6eDUuSPv9M3aSGxWCUhBoKc1S 0.12540630 BTC
1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	100.00000000 BTC		37phUjANXpMhcPnwJdfWFJeUF8gb7mC4HG 0.00096191 BTC
bc1q7x66gw2gal0mne0d4j6l9znx0fwy8sdmur...	0.00877373 BTC		1AqmcLAp7k6dLyN1X876taZ1PYoBDTxx3X 0.06046048 BTC
bc1qhr7rtgw9suj44tts2jgq4hp8gtrpjlrarpkt	0.00844288 BTC		37Pw8Ko97iXqH8epvgfMYnpJMGwtQdi83j 0.04161500 BTC
1NDyJtNTjmwk5xPNhjgAMu4HDHigtobu1s	4.31183454 BTC		bc1q62xuhu04mn92lp2h77990h266x0g8uzf7y... 0.54316389 BTC
bc1qh8t55hpqtdvas2q2v0zrz4z5hlaey9g7e9...	0.00943004 BTC		bc1qenm2vyj3zx03ex54swcxprhtfkn6ytj246ls9z 0.00758266 BTC
			3JtxerEmrDBQbVWvpG5DL8R14cczMSKmf 0.00170000 BTC
			1L5JRGqZdosPGS3BheaeJqLBho5yEQGcgC 0.00852559 BTC
			Load more outputs... (41 remaining)
			156.11445185 BTC

All addresses owned by the same person

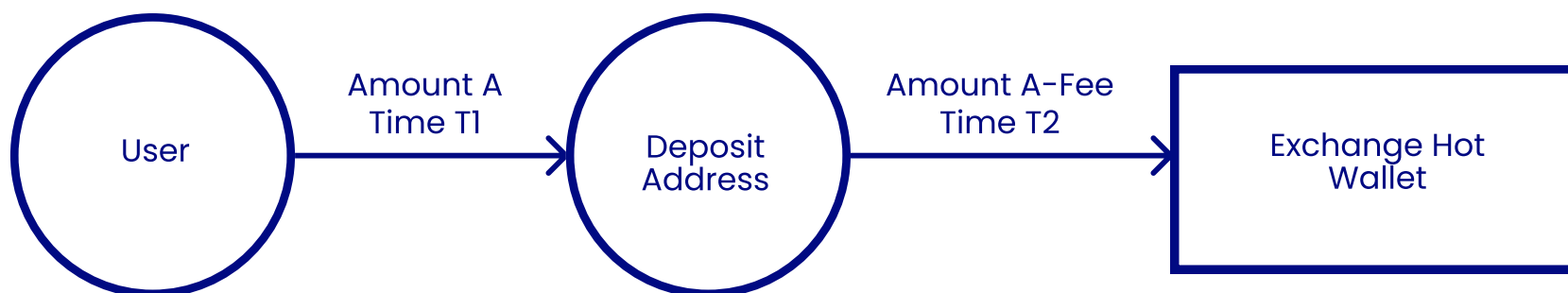
The Common Spend Heuristic empowers us to accurately identify and link addresses owned by the same individual who possesses address X, enhancing our ability to establish comprehensive connections and attribution.

## EVM-centric blockchains

### • Deposit Address

When conducting an examination of an entity, the initiation of a deposit leads to the generation of a distinct deposit address. However, it has been observed that certain entities repeatedly utilize the same deposit address for all their deposits. By tracking the flow of funds originating from these addresses, we can uncover the primary hot wallet of the platform.

Once we have successfully identified a hot wallet, our proprietary algorithm aids in the identification of additional deposit addresses associated with that specific hot wallet with high accuracy.



### • Smart Contract-based Clustering

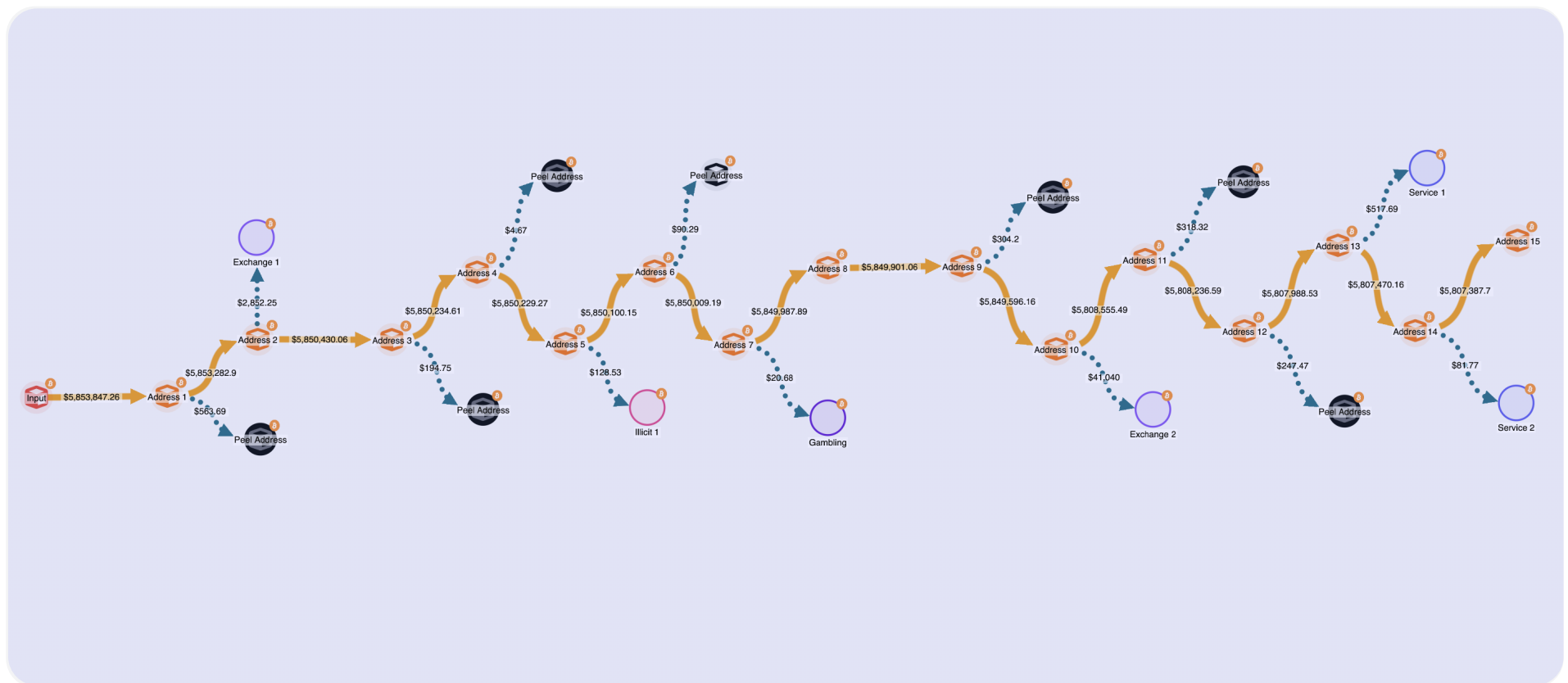
In certain instances, exchanges employ smart contracts in a highly specialized manner. We've discerned distinctive operational patterns in the way these entities function. To discern and organize smart contracts associated with these entities, we employ a clustering approach rooted in smart contract technology.

# Types of clustering based on granularity levels

## • Behavior-based Clustering

Behavioral clustering is the process of deciphering patterns in transactional behavior based on when it occurs and in the structure in which it is executed.

One example of such a behavioral pattern is chain peeling in which illicit actors iteratively break the funds into multiple addresses and ultimately consolidate these peeled funds into a single address. Similarly, phishing addresses exhibit consistent patterns that we've successfully pinpointed and leveraged for clustering.



Visual Depiction of Chain Peeling (Source: Tracker)

## Custom Heuristics

Certain entities can be unequivocally de-anonymized with an accuracy rate of 100% using our cutting-edge custom heuristics.

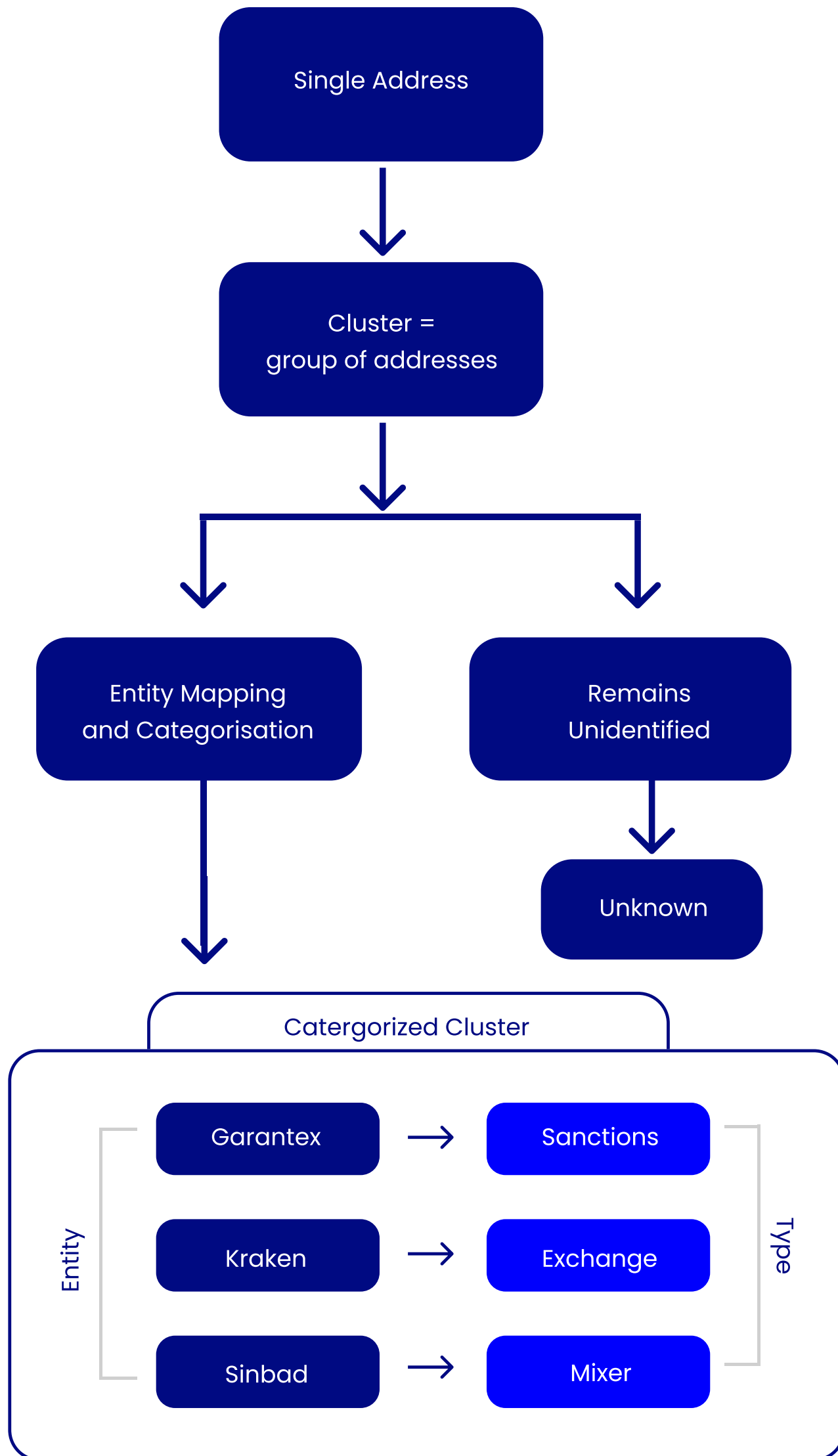


- All Bitmex addresses start with "3BMEX" or include "mex" in their identifier, enabling precise identification and attribution.



- Samurai Wallet Insight: Samurai wallet addresses utilize bech32 format for both input and output, with all outputs consistently totaling 0.05 BTC, facilitating accurate tracking and attribution.

The following diagram details Merkle Science's identification process at a high level:



**Note:**

Entity mapping and categorization is done based on our internal taxonomy. For instance, we use a well-defined taxonomy to classify address annotations.

Addresses and Clusters of Addresses will be classified into one of:

**17+ Entity Types**  
**67+ Entity Sub-Types**

<p><b>Darknet</b></p> <p>ENTITY SUB TYPE</p> <p>Hitman Services</p> <p>DESCRIPTION</p> <p>Entities that provide services in which one party pays a third party (commonly referred to as a hitman) to kill a specific person or group of individuals.</p>	<p><b>Darknet</b></p> <p>ENTITY SUB TYPE</p> <p>Stolen Accounts Vendors</p> <p>DESCRIPTION</p> <p>Entities that operate websites selling stolen account information.</p>
<p><b>Darknet</b></p> <p>ENTITY SUB TYPE</p> <p>Human Trafficking</p>	<p><b>Darknet</b></p> <p>ENTITY SUB TYPE</p> <p>Compromised Credit Cards</p>

**Point and Cluster Tagging**

In some cases, an address may be a part of a cluster of a VASP but at an individual level it can also be associated with another entity/individual/activity based on our attribution.

**For Example- Owner Tag:** Kraken  
**User level:** Scam

*“Operating a scam out of the wallet hosted at an exchange”*

Our customers and partners can suggest clustering preferences for our tools. Once validated, we integrate them into our database.

User Owned Custom Cluster: In the course of investigative processes, users may disclose addresses that are suspected to be associated with specific entity/incident.  
For example: Suspected FTX Hacker

We offer a total of 17 primary categories, each of which is further subdivided into additional subcategories.

## Proof of Reserves

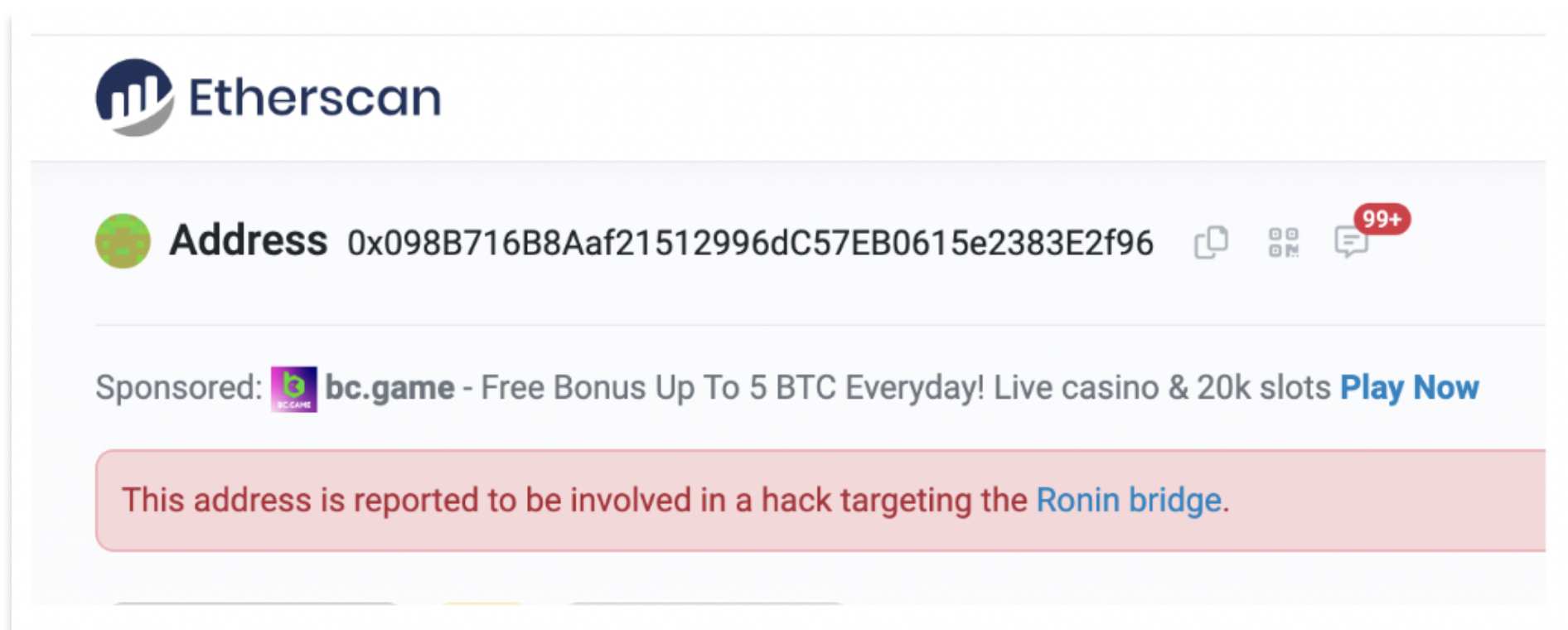
Often, organizations openly disclose their wallet addresses on their publicly accessible websites. Moreover, certain entities furnish proof of reserves as a demonstration of their asset holdings. Proof of Reserves (PoR) is an auditing practice for cryptocurrency companies that can prove their solvency by verifying that the customer assets held by an exchange correspond to the number of assets the company holds in reserve on behalf of its customers. In such instances, we extract these addresses from the public listings and incorporate them into our database.

The reliability of this approach is 100% since these addresses are directly supplied by the entities themselves and are easily verifiable.

## Scraping from Public Blockchain Explorers such as Etherscan

We also employ a data collection approach by sourcing information from public platforms like Etherscan and Tronscan. Notably, blockchain explorers like Etherscan offer annotated address labels, which we find particularly valuable. We maintain a daily streamlined pipeline dedicated to data scraping from selected publicly available sources. However, it's important to acknowledge that not all the sources we extract data from are equally reliable. We place full trust in Etherscan's accuracy, categorizing it as a 99% reliable source. The data scrapped through such sources are verified by our internal team before being it to the database.

Conversely, for other sources, such as those originating from the publications of investigative entities, our trust remains strong, but we exercise a slightly more cautious approach, assigning them an 80% accuracy rating. Data obtained from such sources are thoroughly investigated before being added to our database. All remaining sources, however, are associated with comparatively lower levels of accuracy.



The screenshot displays the Etherscan logo at the top left. Below it, a green circular icon is followed by the text "Address" and the hexadecimal address "0x098B716B8Aaf21512996dC57EB0615e2383E2f96". To the right of the address are icons for copying, QR code, and a notification bubble with "99+". Below the address is a sponsored banner for "bc.game" with the text "bc.game - Free Bonus Up To 5 BTC Everyday! Live casino & 20k slots Play Now". At the bottom, a red warning box contains the text: "This address is reported to be involved in a hack targeting the Ronin bridge."

## **Contributions from LEA and Affiliated Partners**

Through a robust network of partners, clients, and collaborations with Law Enforcement Agencies (LEA), we enhance our attribution precision. Addresses that are sourced from our network of associates are inherently linked to their operations, thus ensuring 100% reliability.

While the LEA data maintains equal credibility, we delay attributions until their official investigations conclude. These addresses are further, methodically assimilated into our internal investigative processes to ensure their accuracy. If necessary, we refer our customers to the designated contact at the LEA for additional information.

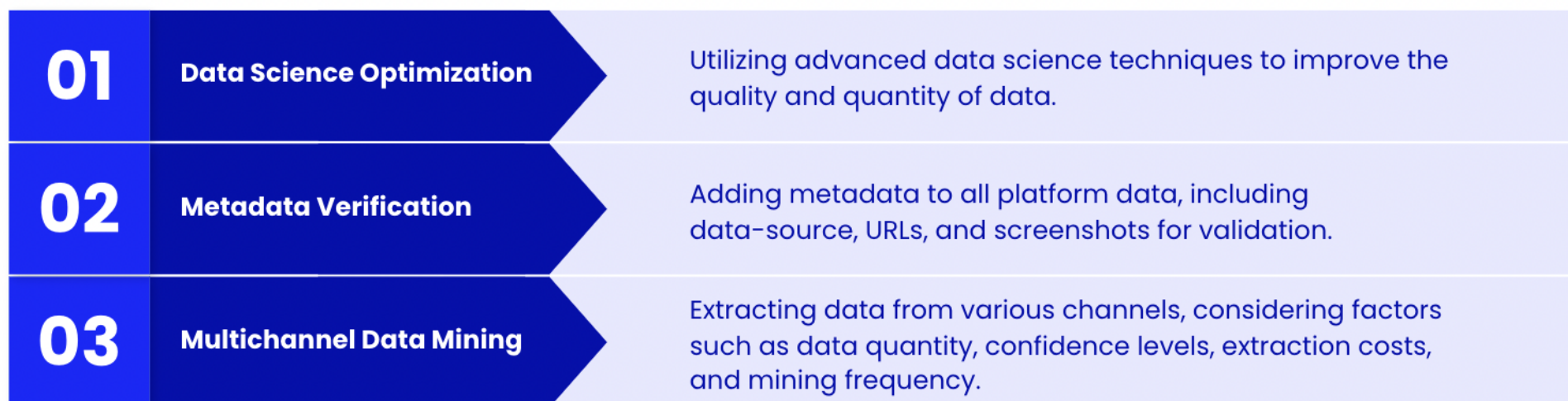
## **OSINT and In-House Investigations**

In our attribution processes, we also leverage Open Source Intelligence (OSINT) to enhance our attribution capabilities. This involves the strategic integration of publicly available data and information from a wide array of sources to optimize our attribution workflows. The inclusion of OSINT significantly enriches the scope and depth of our intelligence, leading to more accurate and robust attribution results.

Our in-house investigations on top of OSINT also reveal numerous addresses associated with a specific entity. Investigations are conducted using proven methods like following the money movement and identifying potential wallets linked to the entities.. We then attribute these specific addresses accordingly.

Beyond the above techniques, we maintain proofs for all the addresses we have collected/identified. This helps our clients and LEAs with case building and putting up strong evidence in the court.

## Assuring Data Integrity Through Our Validation Process



## How do we standardize and ensure quality of data collected from such vastly different streams?

We use a combination of validation rules and human intelligence



## About Merkle Science

Merkle Science provides predictive blockchain risk intelligence and monitoring services that empower compliance teams to prevent illicit cryptocurrency activities and exceed regulatory requirements with confidence. We've raised over \$24M in funding from top-tier investors and are the leading innovator in blockchain analytics. Our solutions power over 100 crypto companies, law enforcement agencies, and financial institutions.

## Contact us

✉ [contact@merklescience.com](mailto:contact@merklescience.com)    🌐 [www.merklescience.com](http://www.merklescience.com)

---

## Follow Us



[merklescience](https://www.linkedin.com/company/merklescience)



[@MerkleScience](https://twitter.com/MerkleScience)



[merklescience](https://www.facebook.com/merklescience)



[Merkle Science](https://www.youtube.com/channel/UC...)