Identification Framework for Cryptocurrency Transaction

Suryani Chang, Nathaniel Tsang Mang Kin
IAME Limited, Mauritius
info@iame.io

Abstract: A decentralised identification framework that would allow cryptocurrency users to verify the identity of each other without having to disclose non-essential confidential information; and to ensure that the users are not involuntarily involved in any illicit activities. Regulatory integration and mainstream adoption depend highly not only on implementation of Know Your Client policy, but also on Anti-Money Laundering and Counter-Terrorism Financing processes, both of which are difficult to conduct when all cryptocurrency addresses are anonymous and transactions are decentralised. The proposed framework would be built in such a way as to be integrated directly with existing blockchains and active cryptocurrency based services to establish an acceptable level of transparency and trust. This whitepaper discusses the problems that are the most critical in the industry and how the framework resolves them.

Keywords. Digital Identity, Document Sharding, Data Minimisation, Link Analysis, Anti Money Laundering, Counter Terrorism Financing.

1. Introduction

Under the original Bitcoin whitepaper[1], it was made intentional by 'Satoshi Nakamoto' that a user's identity remains anonymous. This principle was replicated in all subsequent developments which resulted in the blockchain ecosystem being wholly anonymous: transactions occur between blockchain addresses which have no reliable identification attached to them.

In the absence of any identity linked to blockchain addresses, there has been a lot of regulatory concern that this anonymity is being actively abused in illegal transactions, such as drug trade, extortion, money laundering and potentially terrorism financing. It is from this perspective that we consider identification for cryptocurrency as being the key to regulatory integration, and mainstream adoption.

Identifying a user involves solving the questions of what we know about the user and whether we can verify that the is are who he says he is. However, when identifying transactions, we aim to understand who the user is interacting with and in some cases prevent such transactions from taking place.

Several attempts have been made over the past few years at building identification systems but none have tried to identify cryptocurrency transactions. Instead the focus has been on building centralised platforms to store user information which are now facing scrutiny from privacy laws such as the European General Data Protection Regulation[2] (GDPR). Changes in the law regarding identity data storage will become more severe and within a matter of years, and identification as we know will cease to exist.

Under the IAME framework, the way identity information is stored, shared and validated is completely decentralised, making it perfectly compatible with the nature of blockchain. This whitepaper addresses the construction of the identification framework as follows: section 2, Identity Decentralisation, explains how to construct a decentralised identity; section 3, Data Validation and Sharing, shows how the identity is validated and used; and section 4, Transaction Analysis, demonstrates how the identified addresses can be used to detect illicit activities.

2. Identity Decentralisation

Identity is the first module of the framework. It is built by ascribing attributes such as first name, last name, age, phone number, etc. to particular persons, and proving the authenticity of those attributes. The more attributes a person is ascribed, the more defined is his identity; and the more those attributes are verified, the more genuine is his identity.

Conventional identification requires people to share as much personal information through the transfer of sensitive documentation to counterparties which is a major security to the person sharing the data. This is transforming conventional identification from security systems to a security risks, and is neither scalable nor compatible with the new decentralised ecosystem.

The decentralisation of identity needs to be prioritised, starting with how the identity is constructed, and how the identification data is stored in order to ensure user safety. Within the IAME framework we will explain: how we construct identity, how we shard information for identification and storage on the blockchain, and how we make use of consensus system to achieve a decentralised identity status.

2.1 Constructing an Identity

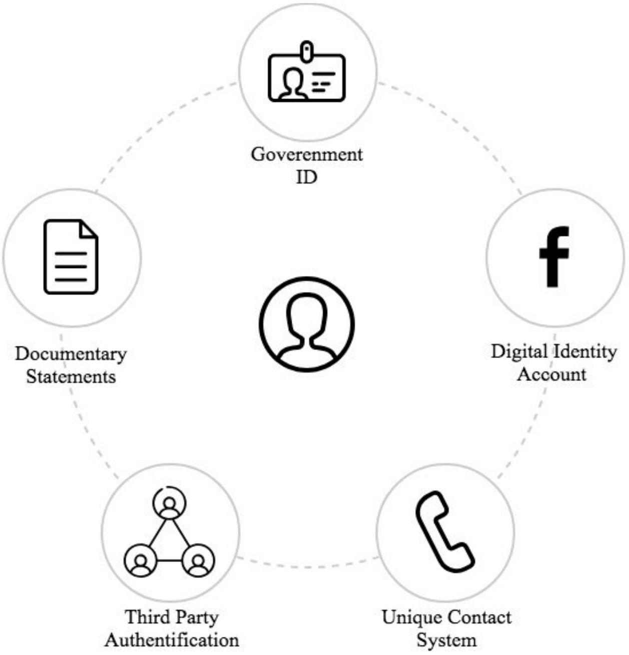


Fig. 1. Sources of Identity

Identities are not the fixed markers, but dynamically constructed[3] from two main sources: orthodox and unorthodox. Orthodox sources have a relative tangibility such as: government issued identification documentations, trusted third-parties statements, and biometrics. Unorthodox sources relate to the identity that can be derived from new technologies such as a phone number, an email or social media accounts. Even if orthodox sources are more trusted than unorthodox ones, both hold significant merit in the establishment of a unique and reliable identity.

To construct an identity, a party submits a statement filled with data about himself or herself, and supports that statement with either orthodox or unorthodox identification sources as supporting documentation. After mapping the statement data to the supporting documentation, a decision is reached on whether the identity of the party can be ascertained.

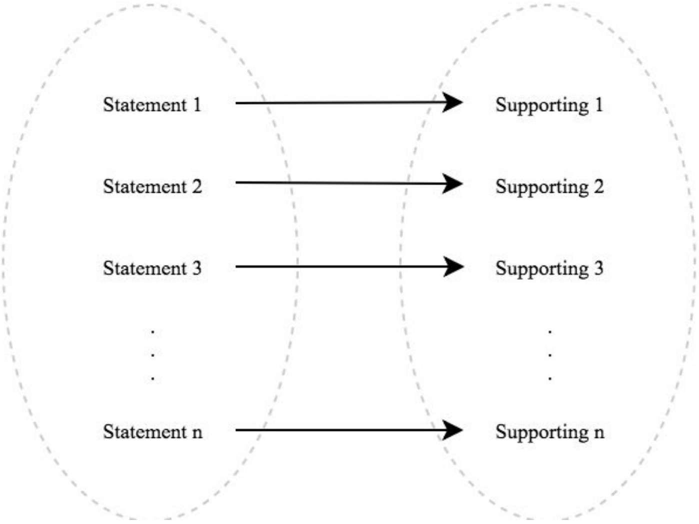


Fig. 2. Identity Information Mapping

The process assumes that: if the party is genuine then all provided statements can be mapped onto supporting documentation. In principle, the more data that can be mapped on more supporting documents, the more difficult it is for a party to falsify its identity. In general identity mapping functions can be categorised as:

1. Statement to string confirmation (e.g alpha numeric attributes);
2. Statement to non-string confirmation (e.g biometrics or photographs);
3. Statement to publicly available information confirmation (e.g public records);
4. Statement to privately available information confirmation (e.g private records); and
5. Statement to government records confirmation (e.g Authority issued document).

Any identification process that can gather a maximum of the above mapping data is ideal for identifiers to guarantee that they are dealing with a genuine person. However, the centralised pooling of those information can be disastrous for the data sharing party in the event of a security breach at the identifier level.

2.2 Document Sharding

Building on the previous identification process, the same mapping functions can be achieved without the party disclosing any non-relevant data to the identifier - by sharding[4] and delegating the mapping functions to unrelated third-party verifiers (TPV). The statement, supporting documentations and functions, would first have to be sharded in such a way that the TPVs who would be confirming shards of data, which on their own, cannot be used by any malicious third party, but the summation of identified shards would constitute a complete identification.

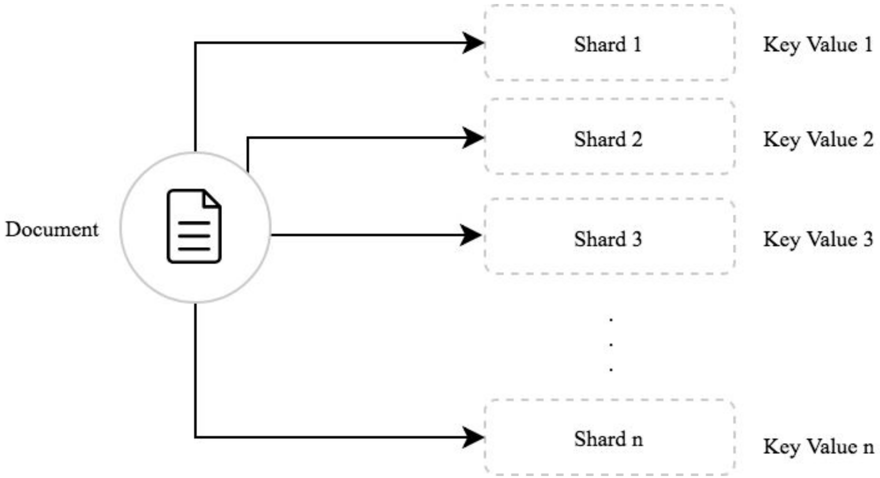


Fig. 3. Document Sharding

Consider a simple scenario whereby a counterparty requires confirmation of 3 string type statements with an identification document as supporting evidence to identify a user:

1. First Name: “John”
2. Last Name: “Doe”
3. Nationality: “British”

In the sharding process, we send the 3 string type statements with their corresponding data shard to 3 different TPVs who would each confirm the content of the string against the data shard. Contrasted with a traditional identification process, in the sharded identification process, each TPVs have a limited amount of information on the user .

However, this sharded identification process can be drawn even further to transform any useful string type statement into unintelligible data, which is how information will be sharded in the IAME framework:

String Type Statement: John

1. Fragment A: “Jo”
2. Fragment B: “oh”
3. Fragment C: “hn”

Using the above process, we have rendered the string type statement, “John”, useless to potential malicious TPVs.

2.3 Consensus Identification

Though sharding is a massive security improvement on wholesome data transfer, it is consensus identification that differentiates IAME from conventional identification systems. By delegating the verification of the shard content to the consensus, we ensure that the wholesome information will never be handled by any centralised entity, thus achieving a truly decentralised identity.

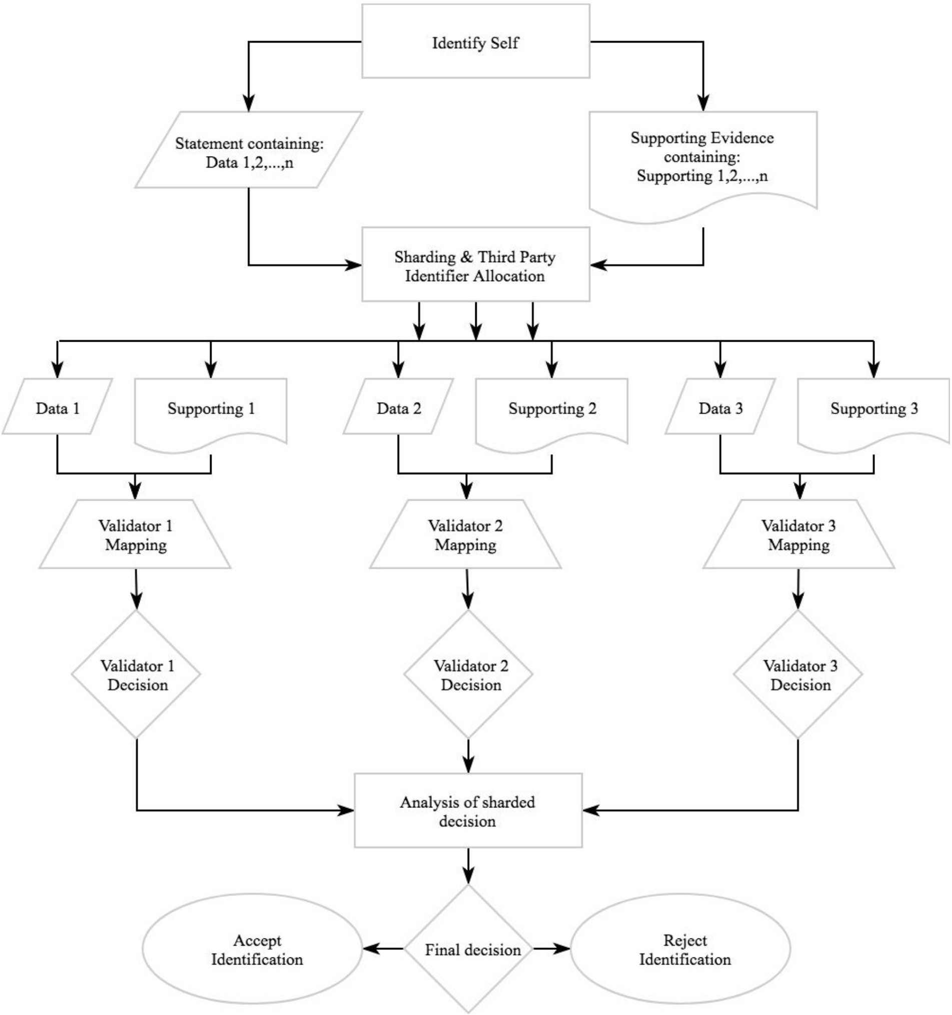
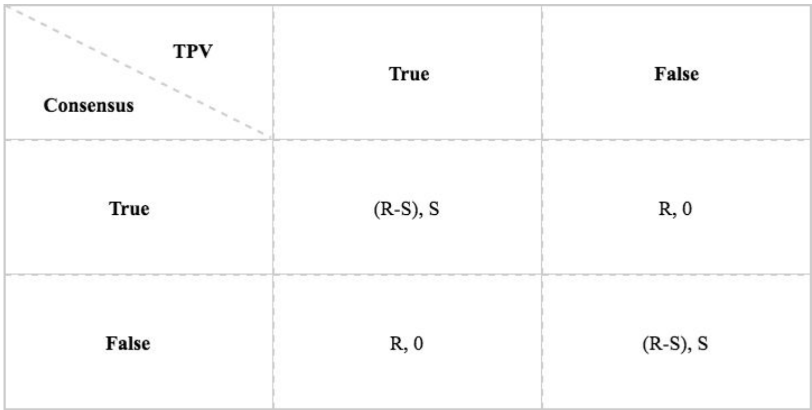


Fig. 4. Consensus Decision Making

The general rationale behind consensus identification is that an identification conducted by a multitude of TPVs is more dependable and less susceptible to the risk of fraud than an identification validation conducted by a single institution or party. However, having a system that runs on a for-profit basis, where the TPVs are "paid" to identify information, creates a natural tendency for participants to game the system and collectively approve a verification process to receive a payout. The solution we propose is a symmetric game model that will encourage truthful verification that is analogous to the Byzantine Fault Tolerance[5] - a Trial Algorithm.



TPV Consensus	True	False
True	(R-S), S	R, 0
False	R, 0	(R-S), S

Fig. 5. Symmetric Game Model

Aspects of the Trial Algorithm:

1. Sharded statement and their supporting document are paired ("Evidence");
2. Evidences are verified by a collection of TPVs ("Jury") in a decentralised process ("Tribunal");
3. The Jury verifies the content of the evidence and gives a decision to the Tribunal; and
4. If there is significant majority on the content of the evidence, the evidence is hashed onto a designated blockchain, else the evidence is sent through a second Tribunal ("Appeal").

The purpose of the Trial Algorithm is to support truthful verification and, most importantly, genuine dissent in the event a significant portion of the TPVs are corrupted. In the proposed technical implementation, it is intended that the tribunals are blindly assigned Tribunals, Appeals, and control Tribunals. Control Tribunals are an allocation of non-matching statements to evidence as a double blind procedure.

3. Data Validation and Sharing

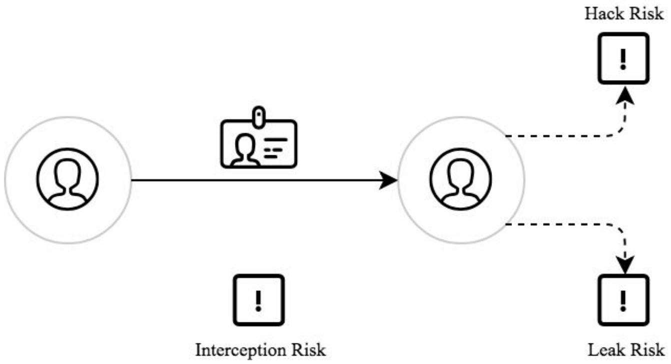


Fig. 6. Risks in Data Sharing

Data validation and sharing is the second module of the framework. How an identity is constructed will determine how data can be validated and shared. Depending on the counterparty, there may be requirement to have that identity validated to ensure the validity of the identity information. This is because identification in itself does not guarantee that a user has not made use of forged information.

To validate identity data, we make use of validators in the form of private enterprise such as Experian, or professionals such as notaries. However the security risks involved in the data validation process are overlooked: interception risk - the information is intercepted during transfer; hack and leak risk - the validators being compromised.

Data sharing in general, whether during the validation process or with transactional counterparties hold significant security risks which we will address. Within the IAME framework we will explain: how we will be using a decentralised application to hold your identity, how user data is shared in the validation process and how to reduce security risk with data minimisation.

3.1 Identity DApp

Most identity systems including the ones that are being built by so called decentralised identity solutions are essentially centralised. This is due to the fact that they either store identity information in their database or, because they have complete control over the identity platform, which gives leeway room for hacking, data leaks, or in the worse case foul play.

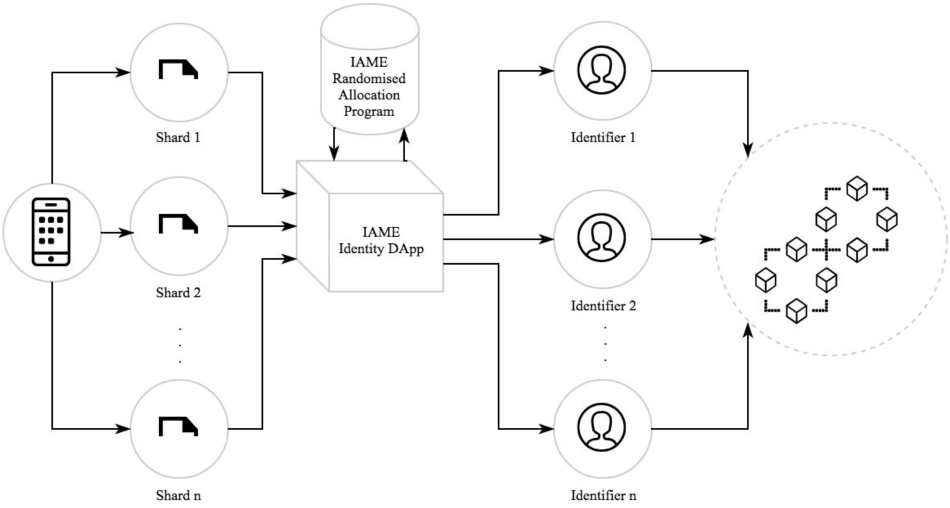


Fig. 7. Identity DApp

At the core of the IAME framework, user identity is held by a decentralised application (DApp) which would be created by the user through his cryptocurrency address. This DApp empowered by the user would be the governing decentralised identity that: distributes the shards to the TPVs and stores the location of the data shards on the designated blockchain.

In essence the existence of the user-controlled identity DApp implies that only the user himself through the DApp would know the location of his identity shards, and have full control with whom he shares his information with. This shift in power is the a main differentiating point that makes identity truly decentralised.

3.2 Validating your Identity

Validation[6] is the process by which identity claims are either cross checked against authoritative sources or asserted by authoritative entities. For example to verify a name, a passport could either be cross checked with the issuing governmental body or be certified by a notary. This process by default requires sensitive information to be transferred by the user which is a security risk, and currently cannot be used to validate identities to cryptocurrency addresses.

In section 2.2 we explained basic sharding, but to make a truly secure validation system, the identity DApp makes use of smart sharding and reconstitution technology, which is an extension of basic sharding. In smart sharding, the identity DApp shards the document only where the data is located, leaving a mold behind such that the document can only be reconstituted if someone has access to the mold and the identity DApp.

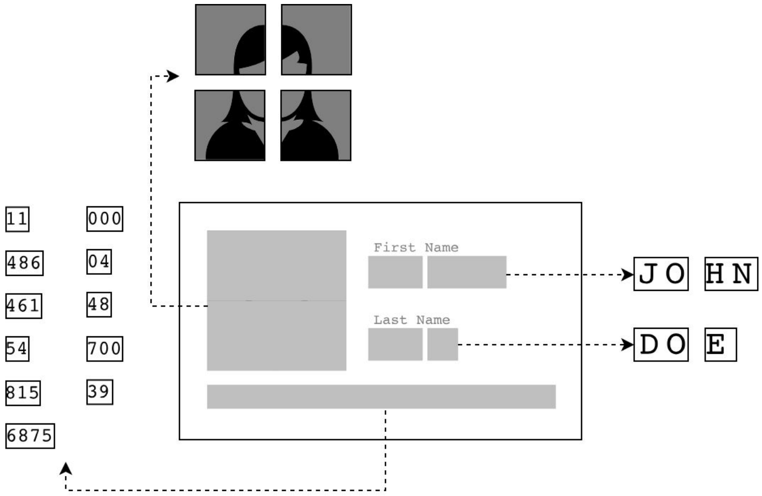


Fig. 8. Smart Shards and Mold

To illustrate how smart sharding and reconstitution would be used, consider a scenario whereby a validator requires a wholesome identity document associated with a cryptocurrency address.

1. The user sends the mold directly to the validator in a traditional system;
2. The user then accesses the identity DApp to authorise the validator to retrieve all the shards associated with the mold.
3. The validator can reconstitute the whole document; and
4. As it is the identity DApp which is enabled the reconstitution, the validator can certify that the validated identity is associated with the user cryptocurrency address.

With IAME, validation is fully secured against interception risk.

3.3 Transacting with Minimised Data

One of the key aspects of the framework is data minimisation[7], whereby the amount of information that is shared by a user to a counterparty is restricted to the strict minimum to reduce the risk of unauthorised access and other impersonation threats. Built on the identity DApp, this function makes use of the results derived from the consensus identification from Section 2.3.

On the IAME framework, data in the identity DApp is kept as both raw data shards and metadata. In the context of data minimisation metadata this would take the form of either structural metadata - that describes how the identity information was constructed; and reference metadata - that describes the status of the data. To illustrate how this would work, let's consider 3 identity data: a first name "John", a last name "Doe", and an address "home"; all of which have been sharded and validated.

Raw Data Shard	**Structural Metadata**	**Reference Metadata**
Shard 1 "Jo"	Data Type: 001 Source: ID Document Length: 2 Position: 1,2	Location: 0x11aa Status: Validated
Shard 2 "hn"	Data Type: 001 Source: ID Document Length: 2 Position: 3,4	Location: 0x22bb Status: Validated
Shard 3 "Do"	Data Type: 002 Source: ID Document Length: 2 Position: 1,2	Location: 0x33cc Status: Validated
Shard 4 "oe"	Data Type: 002 Source: ID Document Length: 2 Position: 2,3	Location: 0x44dd Status: Validated
Shard 5 "Ho"	Data Type: 003 Source: Utility Bill Length: 2 Position: 1,2	Location: 0x55ee Status: Validated

Shard 6 “me”	Data Type: 003 Source: Utility Bill Length: 2 Position: 3,4	Location: 0x66ff Status: Validated

Fig. 9. Metadata Identification

To illustrate how the metadata would be used, consider a scenario whereby a counterparty would require that the first name and last name of a user be identical to his ID document and that the user has a validated address. In that case, the user through the DApp would communicate the structural data of shard 1,2,3 and 4, and only the reference data of shard 5 and 6, without the need to share the shards themselves.

4. Transaction Analysis

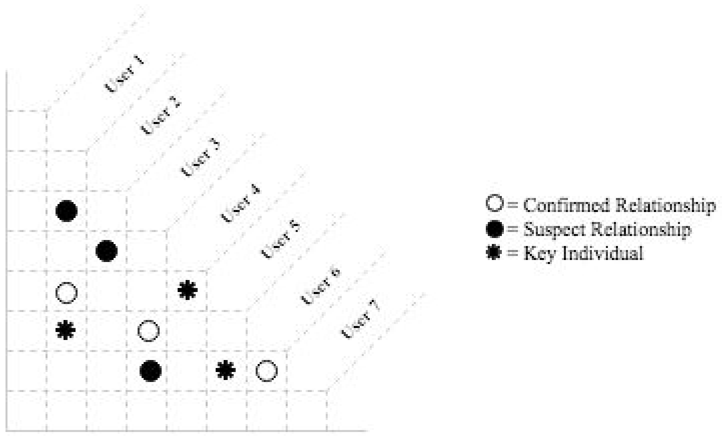


Fig. 10. Association Matrix

Transaction analysis is the third and most important module in the IAME framework, but it can only be achieved once identities are attached to cryptocurrency addresses, and identified transactions enabled. This is because simply analysing transactions between anonymous accounts has no informative value.

By analysing the transactions between identified cryptocurrency addresses, we can access deeper knowledge as to the relationship between them. As a tool it can enable the detection of tainted accounts, e.g accounts that are processing tainted funds or engaging in illicit activities, and isolate them from genuine identified accounts.

Transactional analysis is the most powerful anti money laundering (AML) tool that can be derived from public ledgers but also the most difficult to build. Within the IAME framework we will explain: how link analysis functions, how to conduct multi-dimensional analysis, and how to enact AML with transaction analysis.

4.1 Link Analysis

With public ledgers, all the transactions are accessible to the public, however without the ability to transform those transaction data into a viewable form, the information is useless. Link Analysis[8] (LA) is a data analysis technique used in network theory to evaluate the relationship between nodes in a network.

Under the the IAME framework we analyse complex network relationships using three primary methods:

1. Standard – a simple network that displays all the IAME identified users and how they interact with each other.

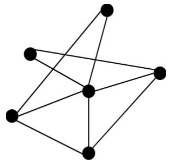


Fig. 11.. Standard Network

2. Structural – users are grouped into clusters based on their structural metadata with each cluster stratified.

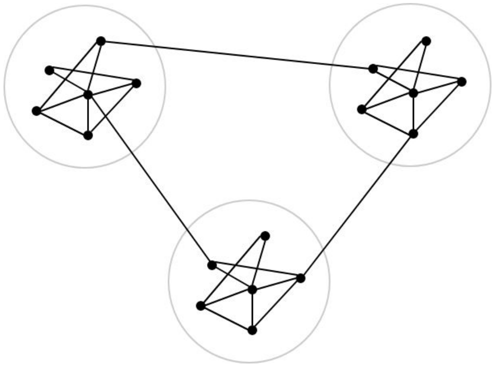


Fig. 12. Structural Network

3. Radial – single users are placed in a center with their connections displayed in a radial tree to establish relationship neighbourhoods.

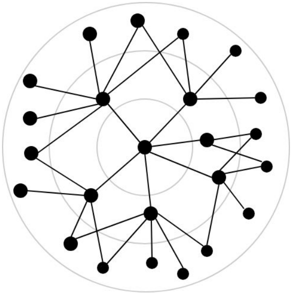


Fig. 13. Radial Network

Using the above we can visualise relationships between users/cryptocurrency addresses, discover hidden relationships and interconnections from seemingly unrelated data.

4.2 Multi-Dimensional Analysis

While LA is an essential tool for transaction analysis, it is bound by a two-dimensional aspect which is the relationship between 2 parties. To look into additional dimensions, we make use of Multi-Dimensional Analysis[9] (MDA) to build multidimensional matrix of relationships. In the proposed framework, this is done by analysing users relationship using metadata values that are built from the Identity DApp through the following data structure:

1. Dimension: structural attribute that is composed of related, hierarchical members.
2. Dimension member: elements in a dimension.
3. Dimension hierarchy: organises dimension members into parent/child relationships.
4. Dimension Title: name by which the dimension is known.
5. Dimension Member Title: the name by which the dimension member is known.
6. Dimension Member Value: an instance of a dimension member.
7. Data Point: the intersection of multiple dimensions.
8. Data: value that resides at a data point.

The advantage of MDA is that it can analyze any kind of distance or similarity matrix. These similarities can represent users' interaction between non-directly related objects, the percentage agreement between transactional parties, or the number of times a user raises an identity query on a counterparty. In general, MDA methods allows the identity framework to mine unobtrusive data and to derive from data underlying dimensions without the user's being compromised.

4.3 Anti Money Laundering and Counter Terrorism Financing

Money laundering is the processing of criminal, “dirty” money to disguise their illicit origin and make them appear legitimate and “clean”. Money Laundering is rarely, if ever, operated by a single persons, business, account, or transaction, but rather by a behavioral pattern[10] occurring over time and involving a set of related parties, which results into AML investigations being generally manual, tedious, time-consuming, and resource-intensive, with a high rate of false positive and inefficiency with voluminous data sets.

Terrorist financing is the process by which terrorists fund their operations in order to perform terrorist acts. Terrorists need financial support to carry out their activities and to achieve their goals. There is little difference between terrorists and other criminals in their abuse of the financial system. While different from money laundering, terrorists often exploit similar weaknesses in the financial system.

Money launderers and terrorist networks are becoming ever more sophisticated in the way they hide their financing. The use of transaction analysis is one of the best ways to be able to visualize the connections between entities, the flow of the funds, and to explain these to law enforcement and regulators.

To be able to resolve the AML/CFT problem in cryptocurrencies, an effective solution is required that can reduce data preparation time, determine detection priority, decrease false positive, and lessen the pressure of manpower, training, and budget. Within the IAME framework, this is achieved using:

1. Decision tree and Bayesian inference to rank suspicious money laundering and terrorist financing cases based on probability computations so as to help AML/CFT analysts focus on the most likely suspects;
2. Employing LA and MDA to identify central members, subgroups, and inter/intra-group interaction patterns in suspect networks;
3. Applying regression and case-based reasoning to uncovering hidden leads and patterns that may prove valuable or timely and predicting prospective trends; and
4. Using support vector machine to deal with high dimensionality heterogeneous data sets.

5. Conclusion

This whitepaper presents the IAME framework as the first generation of truly decentralised identity solution for cryptocurrencies. We have shown how we are building a truly decentralised identity with document sharding and consensus identification which removes most of security risks in conventional identification processes. We have resolved the data validation problem and demonstrated how our identity DApp will empower the user with full control over his decentralised identity, and have provided a clear understanding of what is required to protect users. But most importantly, within the framework we will be able to satisfy the regulatory imperative to be able to conduct AML/CFT on cryptocurrencies. To summarize, the proposed framework will be the key to regulatory integration and mainstream adoption of cryptocurrencies.

Reference

1. Satoshi Namamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System”, 2008.

2. European General Data Protection Regulation (GDPR) (EU) 2016/679, 2016.

3. James D. Fearon, “What is Identity?”, 1999.

4. Rahul Roy, Shard: A Database Design”, 2008.

5. L. Lamport, R. Shostak, M. Pease, “The Byzantine Generals Problem”, 1982.

6. World Bank, “Technical Standards for Digital Identity Systems for Digital Identity”, 2017.

7. T. Antignac, D.Le MétayerPrivacy, “Architectures: Reasoning about Data Minimisation and Integrity”, 2014.

8. A. Ng, A Zheng, M. Jordan, “Link Analysis, Eigenvectors and Stability”, 2001.

9. D. Lemire, “Data Warehousing and OLAP - A Research Oriented Bibliography”, 2007.

10. W. Harper, D Harris, “The Application of Link Analysis to Police Intelligence”, 1975.