用于加密货币交易的识别框架

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

文摘：在QTUM1的基础上建立了一个去中心化的身份标识识别框架，使加密货币用户可以相互验证身份，而不必泄露非必要的机密信息；并确保用户不会非自愿地的情况下参与到任何非法活动中。对于加密货币的监管整合和主流采纳在很大程度上不仅取决于“了解你的客户”政策的实施，还取决于反洗黑钱和反恐融资过程，当所有加密货币地址都是匿名的且交易是去中心化的时後，这两个过程都很难进行。因此拟议框架在建立时将直接与现有的区块链和基于流通的加密货币的服务相结合，以建立足够的透明度及可信性。本白皮书讨论了行业中一些最关键的问题，以及如何通過框架解决它們。

关键词：QTUM、数字身份、文件分片、数据最小化、链接分析、反洗黑钱、反恐融资。

目录

1. 导言

在最初的比特币白皮书2中，中本聪(Satoshi Nakamoto)故意让用户的身份保持匿名。后来的一切发展都遵循了这一原则，并因此导致区块链生态系统完全匿名：也就是说当交易发生在区块链地址之间时，这些地址并没有附上任何可靠的标识。

由于与区块链地址之间没有任何可供识别的身份标识，因此许多监管机构担心，这种匿名性将遭不法分子滥用于进行不法交易如毒品交易、敲诈、洗黑钱和可用于资助恐怖主义等。正是从这个角度出发，我们认为对加密货币的识别是监管整合和获得主流采纳的关键所在。

识别用户不仅仅是解决关于用户的已知信息的问题，还需要能够验证有关用户的信息是否真实无疑。然而，单是识别身份标识并不足以侦查是否涉及非法活动。毕竟进行识别的目的就是为了能够发现可疑的交易，并在有可能的情况下阻止这样的交易发生。

在过去几年中，已经有人几次尝试建立身份标识识别系统，但却没有人从识别加密货币交易中入手。相反，重点一直放在建立集中存储用户信息的平台上，而这样的平台现在正面临着像欧洲通用数据保护条例3(GDPR)这样的隐私法的审查。有关身份数据储存的法律将在几年内变得更加严谨，可能在数年内，我们现在所知道的身份标识识别将不复存在。

在IAME框架下，身份信息的存储、共享和验证方式将完全去中心化，并使之与加密货币完全兼容。本白皮书对身份标识识别框架的构建作了如下论述：其中第2节(去中心化身份)解释了如何构建去中心化身份；第3节(数据验证和共享)说明了身份标识验证和使用的方式；第4节(交易分析)展示了如何利用所识别的地址侦查非法活动。

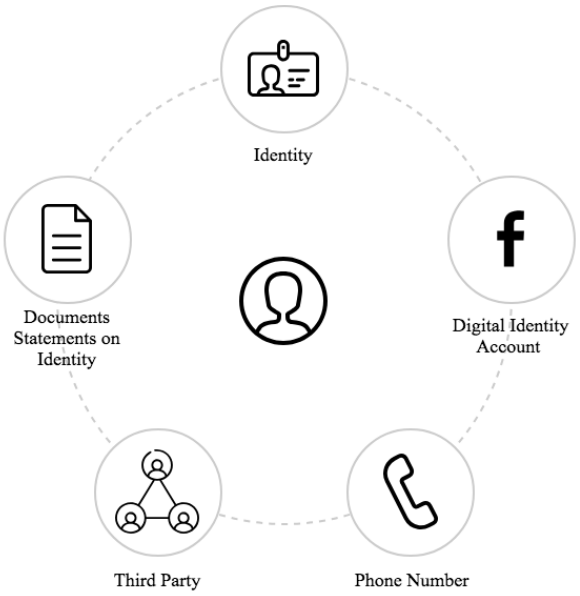
2. 去中心化身份

标识是框架的第一个模块。它是通过将诸如名字、姓氏、年龄、电话号码等属性赋予特定的人，并证明这些属性的真实性来构建的。一个人被赋予的属性越多，他的身份就越明确；这些属性被验证得越多，他的身份就越真实。

传统的身份标识识别要求人们通过向交易方传递敏感文档来共享尽可能多的个人信息，对于共享这些数据的人来说，这些都是需要受到保护的。而且传统的身份标识识别，也将让安全系统转变成安全隐患，这样的方式既不具有可扩展性也无法与新的去中心化生态系统兼容。

在将身份去中心化时需要依据优先秩序，首先是如何构建身份，之后是如何存储身份数据，以确保用户安全。在IAME构建的框架内，我们将论述：我们如何构建身份、如何将信息进行分片用于在区块链上作标识和存储，以及如何利用共识系统来实现去中心化的身份状态。

2.1 构建一个身份



Identity - 身份标识

Digital Identity Account - 数字身份账户

Phone Number - 电话号码

Third Party - 第三方

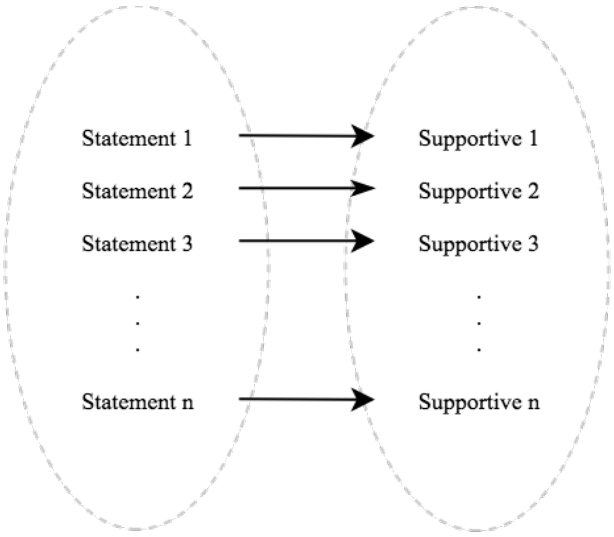
Documents Statements On Identity - 关于身份的文档声明

图表 1. 身份来源

身份不是固定不变的标志，而是由正统和非正统两个主要来源动态建构4 的。

正统来源具有相对的有形性，例如：政府发布的身份证明文件、可信的第三方声明和生物特征识别技术。非正统来源涉及到可以从新技术(如电话号码、电子邮件或社交媒体账户)中获得的身份标识。即使正统来源比非正统来源更值得信赖，但两者在确立独特和可靠的身份方面都具有重要价值。

为了构建身份，当事人提交一份关于他或她本人数据的语句，并以正统或非正统的身份识别来源作为佐证文件支持该语句。在将语句中的数据映射到佐证文件之后，即能够就是否可以确定当事人的身份作出决定。



图表 2.身份信息映射

Statement - 语　　**Supportive** - 佐证

该过程假定：如果当事人是真实的，那么所有提供的语句都可以映射到佐证文档。原则上，能够映射到佐证文档上的数据越多，一方就越难伪造其身份标识。一般来说，身份映射功能可分为以下几类：

1）语句与字符串的确认（例如字母数字属性）；

2）语句与非字符串的确认（例如生物特征数据或照片）；

3）语句与公开可用信息的确认（例如公共档案）；

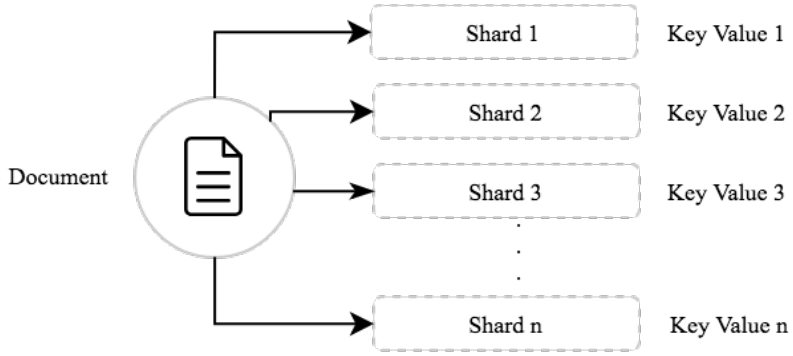
4）语句与私人提供的信息确认（例如私人档案）；和

5）语句与政府记录的确认（例如当局发出的文件）。

任何能够最大限度的收集上述映射数据的识别过程对于标识符来说都是有意义的，以保证它们处理的对象是真实存在的人。但是，在标识符级别出现安全漏洞时，这些通过中心化汇集的信息，对数据共享方来说可能是灾难性的。

2.2 文件分片

在以前的识别过程的基础上，可以在不向标识符公开任何非相关数据的情况下，即通过将映射功能分割并委托给不相关的第三方验证器(TPV)来实现相同的映射功能。语句、佐证文件和功能映射功能首先必须以这样一种方式进行切分，即经由TPV确认的分片讯息无法被任何恶意第三方使用，但已识别分片的总和却可构成完整的身份标识识别。



图表3. 文件进行分片处理

Document - 文档　　**Shard** - 分片　　**Key Value** - 键值

考虑到一个简单的情况，交易方需要确认3个字符串类型的语句，并将标识文件作为标识用户的支持证据：

1. 名字: “John”
2. 姓氏: “Doe”
3. 国籍: “英国”

在分片过程中，我们将3个字符串类型语句及其对应的分片发送给3个不同的TPV，每个TPV只确认相应分片内字符串的内容。与传统的识别过程相比，在分片识别过程中，每个TPV所掌握的用户信息都是有限的。

但是，可以将这一种分片标识过程进一步分割，以便将任何可用的字符串类型语句转换为不可理解的数据，这就是信息在IAME框架中分割的方式：

字符串类型语句：John

1）碎片A：“Jo”

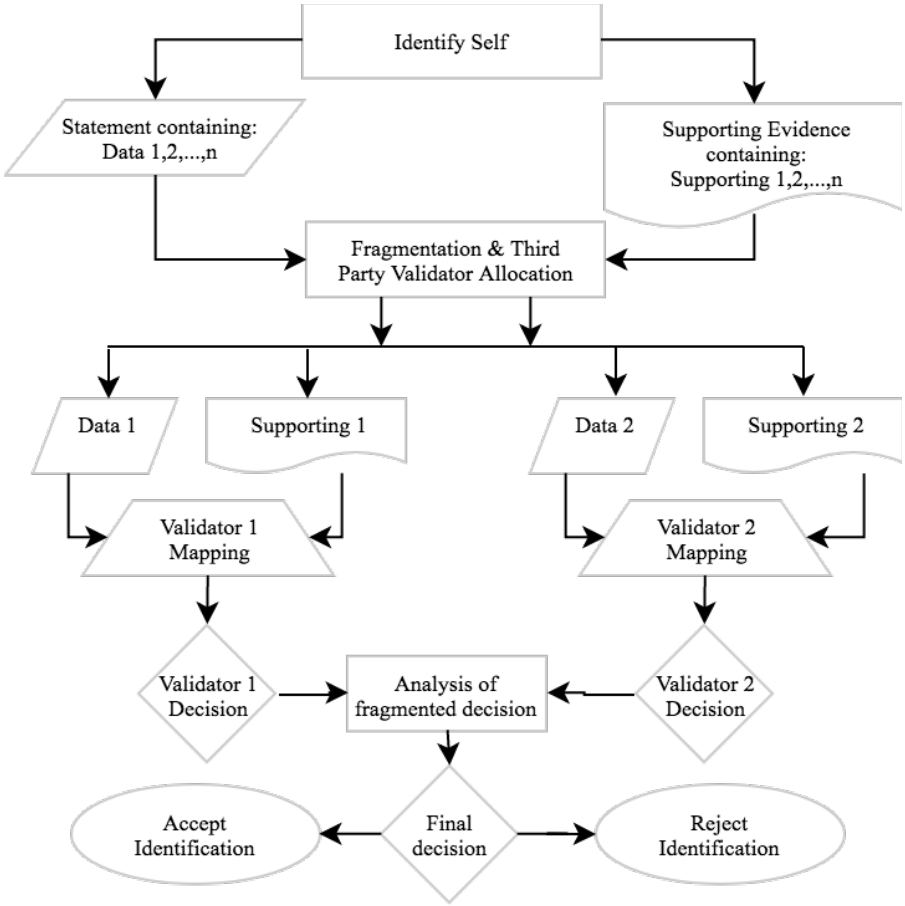
2）碎片B：“hn”

3）碎片C：“oh”

通过上述这样的过程，我们已经让字符串类型语句“John”对任何潜在的恶意第三方验证器都不再有意义。

2.3 识别共识

虽然分片技术对于数据传输带来了有益的大规模安全改进，但将IAME与常规识别系统区分开来的是在身份标识识别上所采用的共识机制。通过将分片内容委托给共识机制确认，我们确保这些有用的信息不会经由任何中心化平台处理，从而实现真正的去中心化身份。



图表 4. 共识决策

Identify Self - 自识别

Statement containing:Data 1,2,...,n - 语句含有：数据 1、2、...、n

Supporting Evidence containing: Supporting 1,2,...,n - 支持证据含有：佐证1、2、...、n

Fragmentation & Third Party Validator Allocation - 碎片化及配置给第三方验证器

Data 1 - 数据 1

Data 2 - 数据 2

Supporting 1 - 佐证 1

Supporting 2 - 佐证 2

Validator 1 Mapping - 验证器1映射

Validator 2 Mapping - 验证器2映射

Validator 1 Decision - 验证器1的决定

Validator 2 Decision - 验证器2的决定

Analysis of fragmented decision - 碎片化决定分析

Accept Identification - 接受身份识别

Reject Identification - 拒绝身份识别

Final Decision - 最终决定

共识识别背后的基本原理是，与单一机构或一方进行的身份标识验证相比，由大量TPV进行的识别更可靠，更不容易受到欺诈风险的影响。然而，拥有一个以营利为基础的系统，即TPV在识别信息后将获得酬劳，这就会产生一个自然趋势，即参与者倾向滥用系统的规则，通过集体批准一个验证过程以获得付款。我们提出的解决方案是一个对称的博弈模型，它将鼓励真实的验证，类似于拜占庭式的容错6这样的一个试验算法。

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

图表5. 对称博弈模型

TPV - 第三方验证器　　**True** - 真实

Consensus - 共识　　**False** - 虚假

试验算法的各个方面：

1. 语句分片及其佐证文档进行配对(“证据”)；

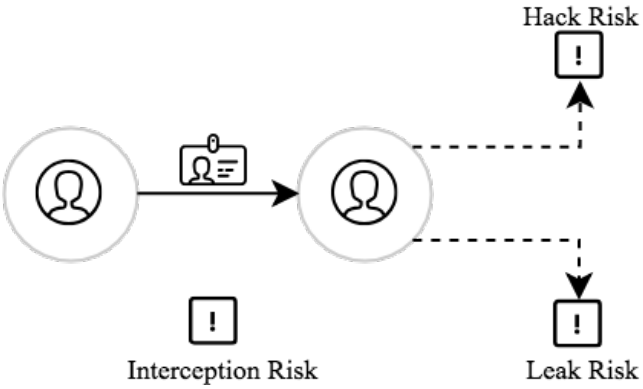
2. 证据由一组TPV(“陪审团”)经由分布式程序(“法庭”)进行验证；

3. 陪审团核实证据的内容，并将决定提交给法庭；以及

4. 如果对证据的验证形成多数共识，则将证据散列到指定的区块链上，否则证据将发往第二法庭（“上诉”）。

试算法的目的是为了支持真实验证过程，最重要的是，可支持真正的异议，特别是发生第三方验证器在验证机制中出现大范围出错的时刻。在我们所提出的技术实施建议中，法庭将随机分配为法庭，上诉庭和控制法庭 。而控制法庭将会把不匹配的语句作为双盲程序分配给证据。

3. 数据验证和共享



图表 **6.** 数据共享的风险

Hack Risk - 被骇风险　　**Interception Risk** - 拦截风险　　**Leak Risk** - 泄漏风险

数据验证和共享是框架的第二个模块。如何构造身份将决定着验证和共享数据的方式。根据对方的情况，可能需要对该身份进行验证，以确保身份信息的正确性。这是因为身份标识识别本身并不能保证用户没有使用伪造信息。

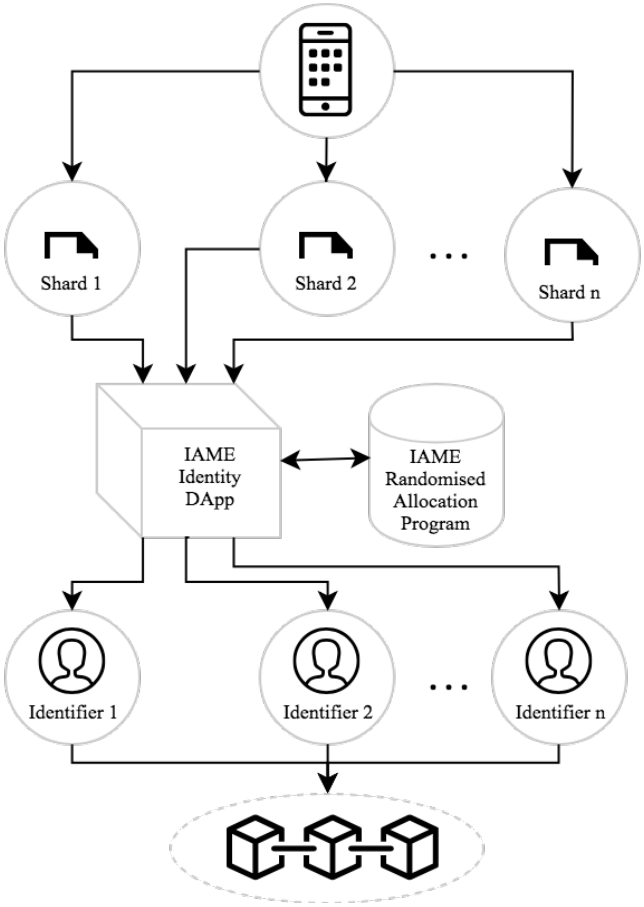
在验证身份数据方面，我们使用私有企业(如Experian)或专业人员(如公证人)

作为验证器。然而，数据验证过程中涉及的安全风险被忽略了：如拦截风险-信息在传输过程中被截获及被骇和泄漏风险-验证器被破坏。

一般来说，无论是在验证过程中还是与交易对手共享数据，都存在着重大的安全风险，而我们将负责处理这些风险。在IAME框架内，我们将解释：我们将如何使用基于QTUM的去中心化应用来保存您的身份，如何在验证过程中共享用户数据，以及如何通过最小化数据来降低安全风险。

3.1 去中心化应用（DApp）身份

大多数身份标识识别系统，包括那些由所谓的分布式身份解决方案构建的系统，基本上都是中心化的。这是因为他们要么将身份信息存储在数据库中，要么对身份拥有完全的控制权，这就为黑客侵入、数据泄漏或更糟糕的情况下的非法行为有了可乘之机。



图表7. 身份标识去中心化应用

Shard - 分片

IAME Identity DApp - IAME身份标识去中心化应用

IAME Randomised Allocation Program - IAME随机分配计划

Identifier - 标识符

在IAME框架的核心中，用户身份标识由一个去中心化应用(DApp)持有，该应用程序将由用户通过他的QTUM地址创建。这个DApp作为用户授权的一系列交互QTUM智能合约构建而成，它将是一个管理的去中心化身份标识：将分片分配给TPV，并将分片数据的位置存储在指定的区块链上。

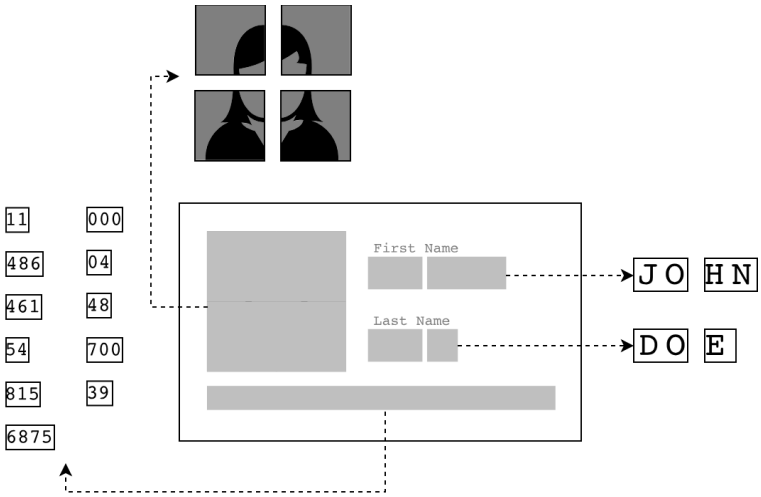
本质上，用户控制的身份标识DApp的存在意味着只有用户自己通过DApp才能知道他的身份标识分片的位置，并有能力完全控制要与谁共享他的信息。

最大的区别是对身份控制权的转移，这也是让身份标识真正做到去中心化的一个关键。

3.2 验证你的身份

验证7是身份声明与权威来源交叉检查或由权威实体断言的过程。例如，为了核实姓名，可以将护照交由签发护照的政府机构交叉核对，也可以由公证人核证。在默认情况下，此过程要求用户传输敏感信息，因此会造成安全风险，这使得这样的方式目前还不能用于验证到加密货币地址中的身份标识。

在第2.2节中，我们解释了基本分片，但是为了建立一个真正安全的验证系统，身份标识DApp使用了智能分片和重构技术，这也是基本分片的扩展。在智能分片中，身份标识DApp仅在数据所在的地方对文档进行切分，并留下一个模具，确保只有在同时能够访问该模具和身份标识DApp时，才能重构该文档。



图表 8. 智能分片与模组

为了说明如何使用智能分片和重构，请考虑这样一个场景：验证器需要一个与加密货币地址相关联的安全的身份文档。

1. 在传统系统中，用户将模具直接发送给验证器；
2. 然后，用户访问身份标识DApp，以授权验证器检索与模具相关联的所有分片。
3. 验证器可以重新构建整个文档；以及
4. 由于重构过程中需要通过身份标识DApp，验证器可以证明验证的身份标识与用户加密货币地址相关联。

有了IAME，验证就可以确保安全无虞并完全避免拦截风险。

3.3 以最小化数据进行交易

该框架的一个关键方面是数据最小化8，即用户与对手方共享的信息量被限制在严格的最低限度内，以减少未经授权的访问和其他冒名风险带来的威胁。

在身份标识DApp的基础上，该功能利用了第2.3节中识别共识所得出的结果。

在IAME框架中，身份标识DApp中的数据作为原始数据分片和元数据保存。

在数据最小化元数据的上下文中，这将任一采取结构元数据(描述身份信息是如何构造的)和引用元数据(描述数据的状态)的形式进行。为了说明这是如何工作的，让我们考虑三个身份数据：即名字“John”、姓氏“Doe”和地址“home”；所有这些数据都被分片及已接受验证。

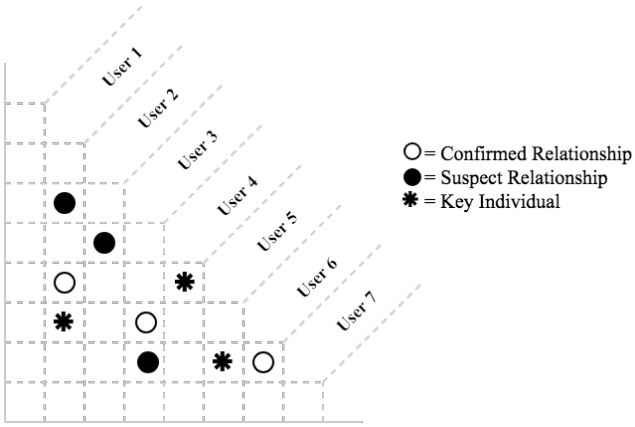
原始数据分片	结构元数据	引用元数据
分片 1 **“Jo”**	数据类型: 001 来源: 身份证件 长度: 2 方位: 1,2	位置: 0x11aa 状态: 已验证
分片 2 **“hn”**	数据类型: 001 来源: 身份证件 长度: 2 方位: 3,4	位置: 0x22bb 状态: 已验证
分片 3 **“Do”**	数据类型: 002 来源:身份证件 长度: 2 方位: 1,2	位置: 0x33cc 状态: 已验证
分片 4 **“oe”**	数据类型: 002 来源:身份证件 长度: 2 方位: 2,3	位置: 0x44dd 状态: 已验证

分片 5 **“Ho”**	数据类型: 003 来源: 公用事业账单 长度: 2 方位: 1,2	位置: 0x55ee 状态: 已验证
分片 6 **“me”**	数据类型: 003 来源: 公用事业账单 长度: 2 方位: 3,4	位置: 0x66ff 状态: 已验证

图表 9. 元数据识别

为了说明如何使用元数据，考虑这样一种场景：交易方要求用户的名字和姓氏与其身份证件相同，并且用户的地址必须经过验证。在这种情况下，用户将通过DApp传输分片1、2、3和4的结构元数据，并且只传输分片5和分片6的引用元数据，而不需要共享分片本身。

4. 交易分析



User - 用户

Confirmed Relationship - 确证关系

Suspect Relationship - 可疑关系

Key Individual - 关键人物

图表10.关联矩阵

在IAME框架中，交易分析是第三个也是最重要的模块，但只有将身份标识附加到加密货币地址，并启用了被识别的交易后才能实现。这是因为简单地分析匿名帐户之间的交易没有信息价值。

通过分析已识别的加密货币地址之间的交易，我们可以更深入地了解它们之间的关系。作为一种工具，它能够发现那些有问题的账户，例如那些用于处理问题资金或从事非法活动的账户，并将它们与那些真实的已确认账户隔离开来。

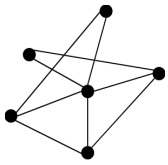
交易分析是最强大的反洗钱(AML)工具，它可以从公共账本中派生出来，但也是最难构建的。在IAME框架内，我们将解释：链接分析是如何工作的，如何进行多维分析，以及如何通过交易分析制定AML。

4.1 链接分析

有了公共账本，公众因此得以访问所有的交易，但是如果没有能力将这些交易数据转换成可查看的形式，这些信息就毫无意义。链接分析(LA)是网络理论中用来评估网络中节点之间关系的一种数据分析技术。

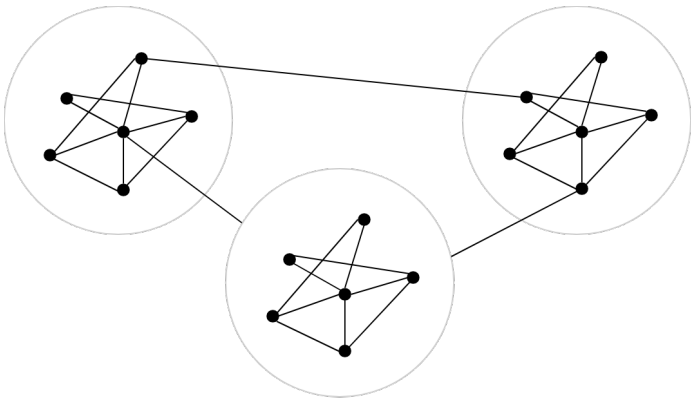
在IAME框架下，我们使用三种主要方法来分析复杂的网络关系：

1. 标准 – 一个简单的网络，显示所有IAME已识别的用户和他们之间如何进行互动。



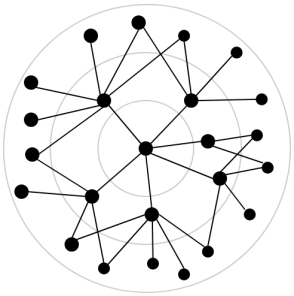
图表11.标准网络

2. 结构 – 根据结构元数据将用户分组为集群，并对每个集群进行分层。



图表12. 结构网络

3. 径向 – 单个用户被放置在一个中心，他们的连接显示在一个放射状的树中，以建立一个社区关系。



图表13. 径向网络

使用上述方法，我们可以将用户/加密货币地址之间的关系变得以可视化方式呈现，从而在看似不相关的数据中发现隐藏的关系和相互关联处。

4.2 多维分析

虽然LA是交易分析的重要工具，但它却受到两个当事人之间的关系这一个二维方面所约束。为了研究更多的维度，我们利用多维分析10(MDA)来建立多维关系矩阵。在拟议的框架中，这是通过身份标识DApp利用以下结构元数据值，以分析用户之间的关系。

1. 维度：由相关的、层次化的成员组成的结构属性。
2. 维度成员：维度中的元素。.
3. 维度层次结构：将维度成员组织为父母/子女关系。
4. 维度标题：已知维度的名称。.
5. 维度成员标题：已知维度成员的名称。
6. 维度成员值：维度成员的实例。
7. 数据点：多个维度的交集。
8. 数据：驻留在数据点的值。

MDA的优点在于它可以分析任何种类的距离或相似矩阵。这些相似之处可以表示用户在非直接相关对象之间进行的互动、交易方之间的百分比协议，或者用户向交易方提出身份标识查询的次数。通常，MDA方法允许身份标识框架挖掘不显眼的数据，并在不损害用户利益的情况下，从底层维度派生数据。

4.3 反洗黑钱和反恐融资

洗黑钱是对犯罪、“肮脏”钱的处理，以掩盖其非法来源，使其显得合法和“干净”。洗黑钱很少通过单一个体、企业、账户或交易进行的，而是通过一种行为模式11进行的，这种行为模式随着时间的推移而发生，并涉及一组相关方，这导致反洗黑钱调查通常涉及大量人力、调查过程繁琐、需要大量时间和耗费大量资源，并且假阳性率（误报）很高，而且缺乏大量数据集。

恐怖主义融资是恐怖分子为其行动获得资金以实施恐怖行为的过程。恐怖分子需要财政支持来开展活动和实现其目标。在滥用金融系统方面恐怖分子和其他罪犯几乎没有区别。虽然有别于洗黑钱，但恐怖分子同样利用金融系统中可被滥用与洗黑钱的的漏洞。

洗钱者和恐怖分子的网络在他们隐藏资金的方式上越来越老练了。交易分析

是能够以可视化方式呈现的，用于向执法人员及监管机构解释实体之间的联系及资金流动的方式的最佳方式之一。

为了能够解决加密货币中的AML/CFT问题，需要一种有效的解决方案，这种方案必须可以减少数据准备时间，确定检测优先级，减少误报、人力、培训和预算的压力。在IAME框架内，可以通过以下方式实现：

1. 利用决策树和贝叶斯推理算法，基于概率计算对可疑洗钱和恐怖融资案件进行排序，以帮助AML/CFT分析人员监控最可能的嫌疑人;

2. 使用LA和MDA来识别可疑网络中的中心成员、子组以及不同组别及组内的互动模式；

3. 运用回归和基于案例的推理，发现可能被证明有价值或及时的潜在线索和模式，并预测未来趋势;和

4. 利用支持向量机处理高维异构数据集。

5. 结论

本白皮书介绍了IAME框架，它是构建在QTUM上的第一代真正去中心化身份的加密货币解决方案。我们已经展示了如何通过文件分片和识别共识构建一个真正去中心化身份的方式，从而消除了传统身份识别过程中的大部分安全风险。我们已经解决了数据验证问题，并演示了基于QTUM的身份标识DApp将如何赋予用户对其去中心化身份进行完全控制，并就如何保护用户作出清晰的说明。但最重要的是，在这个框架内，我们将能够满足监管要求，能够对加密货币进行反洗钱操作。总而言之，拟议的框架将是监管整合和主流采用加密货币的关键。

引文出处

1. Patrick Dai, Neil Mahi, Jordan Earls, Alex Norta, “QTUM: Smart-Contract Value-Transfer Protocols on a Distributed Mobile Application Platform”, 2017.

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

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

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

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

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

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

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

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

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

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