COLENDI TECHNICAL PAPER

Abstract

This paper serves to provide a high-level overview of the proposed technical architecture for Colendi's decentralized credit scoring protocol and micro-credit platform. The traditional way to evaluate the creditworthiness of an individual is to make a prediction over one's financial picture which involves income, expenditure, and credit history. Credit scoring institutions are mostly centralized third-party systems that rely on obsolete techniques and technologies. However, advances in machine learning and the opportunity to compute big data has led us to the evaluation of creditworthiness through social media, mobile phone data, connections, interactions, actions, experiences, and many other relevant data that may serve to build a financial passport for redefining creditworthiness. We propose a model that is composed of a decentralized identity and credit scoring protocol which will be complemented by a decentralized microcredit platform as a proof of concept. This document does not attempt to go into minute details regarding the implementation, as there are likely to be changes due to technical limitations and continuously updated timelines.

I. Overview

At a high level, Colendi Protocol will consist of three core components: a form of identity management, a mechanism for collecting/storing data about users, and a means of generating a credit score/lending decision based on users' data. In order to reduce development costs and gain free network advantages, many of the distributed aspects of these components have been built off of external ecosystems (Figure 1). For identity, we use **ERC-725**, an existing decentralized identity standard with its complementary identity claim standard, namely **ERC-735**. Storing users' personal data is handled via a **Secure Object Storage(SOS)**, facilitated by Colendi. Processing this personal data is executed through the **Secure Computation Environment(SCE)**, based on multiparty computation.

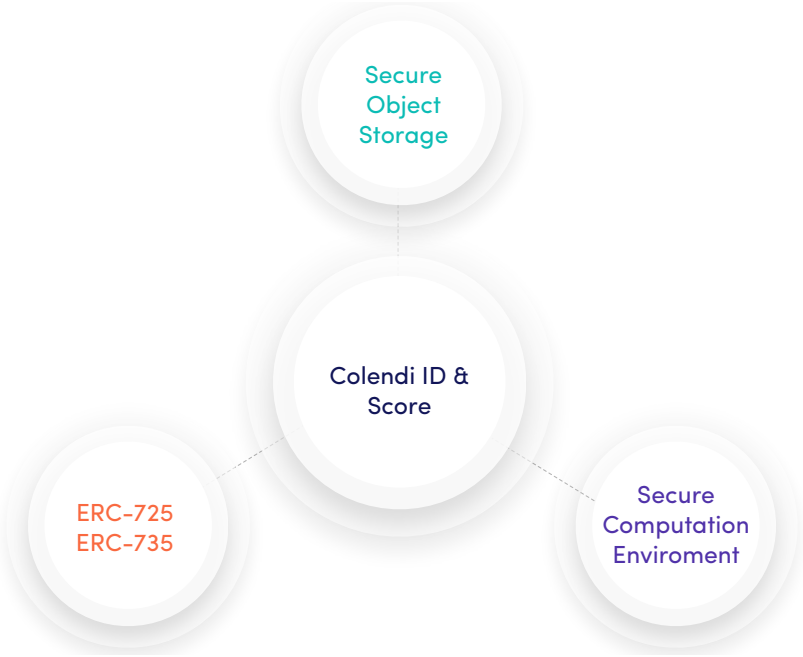

Fig. 1. Colendi Stack with external ecosystems

End-users (*i.e. potential borrowers*) will use Colendi via a mobile application which will facilitate all their interactions with the network. This includes performing basic tasks such as creating/managing identity, key management using *ERC-725*'s proxy contract, collecting/uploading data available on their devices, providing social media data with OAuth, and signing transactions to approve various requests. In addition to these basic tasks, end users will also be able to browse through merchants in Colendi Network and given the option to do so, apply for microcredit.

Collecting and storing data about users is based on two actors, namely **data integrators** and **data partners**. Data integrators are small financial technology firms that will use Colendi's tools to integrate data partners and merchants into Colendi Network. Data partners *(i.e. merchants, telcos, banks, etc.)* that have completed their integration with Colendi Protocol, will be allowed to submit data about users via an application provided by Colendi. This data will be either encrypted with users' public keys or anonymized; and may only properly take place in the Colendi Network after checking existing user approval with data partners. In the early stages of the application, users are required to agree with terms of use to grant permission to Colendi for utilizing necessary user data from trusted data partners. The encryption/anonymization process is handled via Colendi client-side SDK and publishing encrypted/anonymized data to the network is entirely handled via Colendi services.

Potential lenders will be allowed to request a credit score to be run on potential borrowers (*end-users*), which will leverage existing data and a selected **scoring algorithm**, implemented in a Secure Computation Environment, which will allow for a lending decision to be reached. As our Secure Computation Environment functions properly, neither the lenders nor the nodes of the environment in which the scores are calculated, should have access to any private financial or social media data. Lenders will only be granted access to borrowers' information collected by third-party KYC providers integrating with Colendi ID. The chosen **scoring algorithm** may be a first-party implementation by Colendi or a third-party implementation *(created for more task-specific scores/decisions)*. These third-party scoring algorithm developers may be incentivized by receiving a fee each time their algorithm is run for a score query. Colendi will provide a service to leverage existing data for a third-party scoring algorithm for future expansions. Development of these algorithms will be transparent/open-source, otherwise ensuring reliability and security of scoring algorithms would be hard. Colendi will provide a whitelist of scoring algorithms stored in a Colendi-managed smart contract.

II. COLENDI ID : SELF SOVEREIGN IDENTITY

Globally, about 1.7 billion adults remain unbanked without any access to financial instruments [1]. Unfortunately, these people are invisible to mainstream financial systems. The solution Colendi offers to tackle this problem is creating a financial passport that will enable the inclusion of this population in the financial ecosystem that is being disrupted to a new paradigm with the help of blockchain technology. Therefore, we are introducing **Colendi ID**, which is a self-sovereign identity that is completely controlled and managed by the owner. Colendi ID is tightly coupled with our scoring protocol and a major component of our credit score evaluation of unbanked and underbanked people. There are a few key features which an identity solution should demonstrate in order to best fit with Colendi's mission and requirements:

- Provide a decentralized and trustless method of reliably identifying a user's real-world identity.
- Ensure that the users exercise sovereignty on their IDs, not a centralized authority.
- Be sufficiently flexible/expandable to allow for future modifications to the Colendi Protocol.
- Allow a standard and easy implementation across other services that want to utilize Colendi Network, reduce difficulty for users to onboard, and create an additional source of data about a user's credit-worthiness.

In addition to Colendi's requirements for identity, providing a global and borderless identity demands a unique identifier on a distributed ledger, followed by a standard. One of the most common standards for a unique identifier on the Internet was drafted in 2005 with *UUIDs* which are specified in RFC4122 [2]. Such a standard requires implementing within a resource which enables the authentication of an entity in a secure manner. **Decentralized Identifier (DID)**, which is a fundamental building block for *self-sovereign* digital identity, is fully under the control of the DID subject, independent from any centralized registry, identity provider, or certificate authority [3]. DIDs are designed to work with different blockchains, so providing interoperability. To put it simply, it is integral for Colendi ID to implement an identity in compliance with DID specifications.

A combination of **ERC-725** and **ERC-735** appears to be an ideal fit for these requirements and being treated as valid DIDs. By piggybacking off existing standards, we avoid reinventing the wheel *(saving on development time)* in addition to gaining the network effect of the proposals' following. The network effect is an essential factor in forming a reliable identity: the more institutions that rely on and contribute to a form of identification, the more trustworthy it becomes. Colendi will benefit from the open nature of *ERC-725/ERC-735* by using external endorsements on a given identity (especially from trusted authorities) both as a form of identification and a measure of creditworthiness in itself. Along with *ERC-725* and *ERC-735*, off-chain identity would also be included as a component of Colendi ID. Thus, we are able to make claims about users without revealing their sensitive data to third-party applications which have integrated into our identity service.

Integrating *ERC-725/ERC-735* into the network design is relatively straightforward. Upon registration in the Colendi App, the user will be prompted to either create an identity or link their existing identity. Once identity is created/selected, the application will call the *addKey* method on the identity smart contract with the value consisting of a generated public key of Secure Object Storage (*with the private key remaining on the user's mobile device and being stored on secure key management protocol at Secure Computation Environment*). This *ERC-725* identity will then serve as an identifier on Colendi's network. Once this attribute is set, any data partner looking to submit input about a user will be able to inquire about their identity address which in turn will provide them with a public key. They may then use it with Colendi's SDK to encrypt and store data on the network.

Other actors on Colendi Network may also leverage *ERC-725* to provide a form of identity. Data partners, data integrators, potential lenders, merchants, and third-party scoring algorithm developers can create their respective identities to represent themselves while interacting with users. With the claim holder contract *(ERC-735)*, users can obtain a claim approved by the KYC providers stating that they have completed the KYC phase. The lenders may define some KYC providers as trusted issuers if they prefer. Thus, lenders can only lend to the users who have approved claims from these providers.

III. SECURE STORAGE AND COMPUTATION

According to *General Data Protection Regulation(GDPR)*, storing personally identifiable information *(PII)* in public blockchains like Ethereum is not permitted. At Colendi, we have serious financial data collected from our integrated partners and personal data - including credentials. Keeping this significant regulation in mind, we are implementing storage and secure computation mechanism which is entirely GDPR-compliant. The storage mechanism is a very common workaround solution. Personal data is stored off-chain and references to this data are kept on the blockchain or on a Secure Computation Environment. It is important to note that off-chain storage solution in this workaround may be centralized or decentralized, but we have chosen to side with decentralization. We also have to guarantee that there will not be any information leaks while making computations on data. Thus, we will present two external protocols: one for storing data and another for performing secure and reliable computations on this data, *Secure Object Storage* and *Secure Computation Environment*, respectively.

A. Secure Object Storage (SOS)

The storage in our protocol must be scalable, reliable and also must guarantee zero downtime and full data integrity. Current blockchain infrastructures do not enable the storage of large amounts of data in terms of scalability, even though they can be regarded as reliable. Providing a solution built on top of centralized databases or file systems for the sake of privacy may produce vulnerabilities that may result in loss or breach of data. From the perspective of Colendi, decentralized storage complements the very nature of blockchain. Therefore, we have built our protocol on top of blockchain and decentralized storage to achieve complete decentralization.

There are comprehensive attempts to create a decentralized storage network. i.e. Swarm, IPFS, Storj, BigChainDB. Most of these storage platforms implement content-based addressing such as Distributed Hash Tables and content-addressed chunk store. These platforms are mostly seen as file storage systems whereas BigChainDB is a database which is an implementation of NoSQL that lets us query entries in a decentralized manner. Even if it's possible to store a file in BigChainDB network, Colendi stores structured data as encrypted, and so it's not possible to leverage BigchainDB's indexed query mechanism. Thus, we have decided to use Storj among various decentralized storage systems for the following reasons:

- Incentives to nodes in the Storj network to provide availability of stored data
- Continuous project development
- Allowing Colendi to facilitate user interaction with Storj network
- Ability to store confidential or explicit data

The Colendi Network will leverage Storj for safely storing user data in a decentralized manner. In order to reduce complexity for end users (*i.e. borrowers, merchants, lenders)*, Colendi will provide a Storj Bridge node which clients will use to facilitate storage calls. While Bridge may be seen as a point of centralization, it is important to note that it will have no access to keys and does not store any data. Colendi protects the privacy of the client and gives them complete control over access to their data while delegating the responsibility of keeping files available on the network to Bridge [4]. In addition to facilitating interactions between Colendi clients and the network, Bridge will also be tasked with providing funding for data storage.

```
{ "type" : "telco",
  "data" : [
          {
          "timestamp": "1533296896",
          "details": {
             "cellularUsage": "4096",
             "callTime" : [
                     {"incomingCall" : "19021"},
                     {"outgoingCall" : "20012"}
                ]}
        }]
}
```

Listing 1: Sample JSON object collected from a telco

Colendi's SDK should provide storage-related tasks for users by offering a set of functions that rules out the necessity of educating the user about underlying technologies. For example, when encrypting and submitting user data, the SDK will expose a simple function that will take *the JSON object of user data (Listing 1)*, *the address of a user's identity*, and *the data partner's private key*. The function fetches the Storj public key associated with the identity. The JSON object is encrypted with this public key, and it is then signed with the provider's private key. SDK follows by submitting the blob to Storj via Colendi Bridge and appends a reference of the encrypted data blob to the user's list of data references in the off-chain storage of Colendi ID. The activity diagram of this progress is shown in Figure 2.

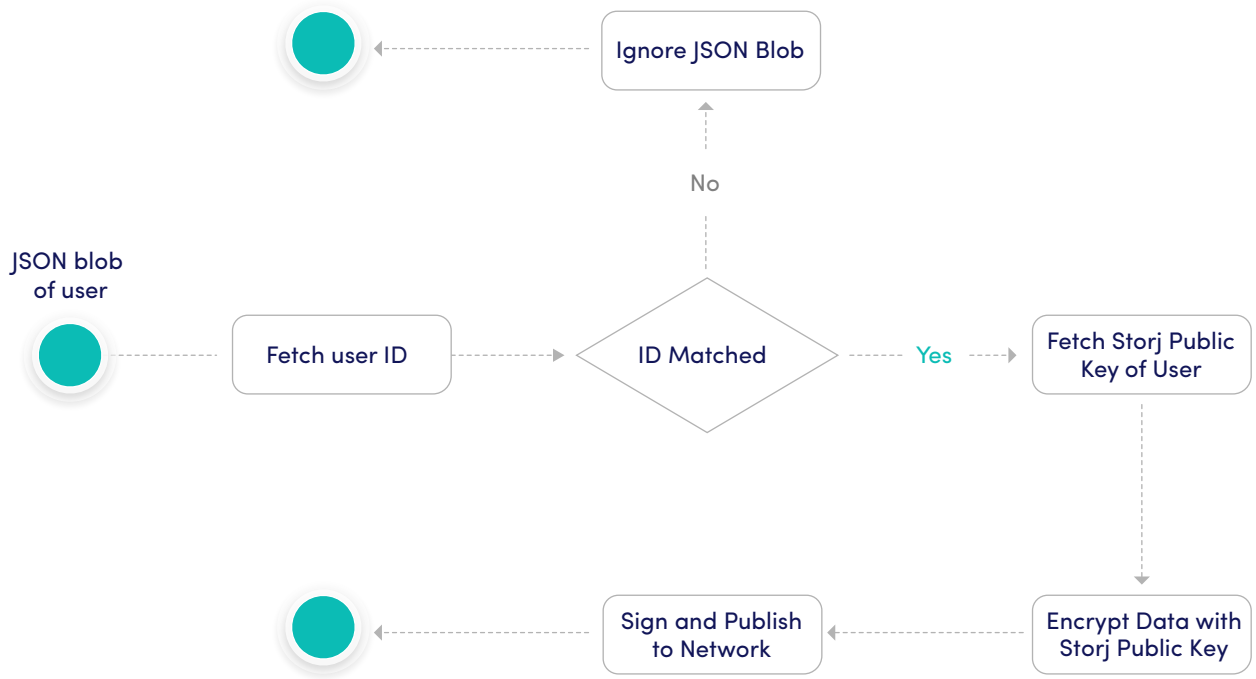

Fig. 2. Signing an encrypted JSON object guarantees data origin authentication.

B. Secure Computation Environment (SCE)

Vitalik Buterin, the co-founder of Ethereum, asserted in his blog-post that *"...it's harder to create a 'holy grail' technology which allows users to do absolutely everything that they can do right now on a blockchain."* [5]. As all transactions on the Ethereum Network are public and costly, we cannot compute the credit score of users on-chain. Yet, we have to preserve user privacy in the Colendi Network. Zero Knowledge Proof *(ZKP)*, Homomorphic Encryption, secure Multi Party Computation *(sMPC)* and Trusted Execution Environment *(TEE)* are all outcomes of the ongoing debate around privacy-preserving applications, and some of them have various implementations. Among these proposed solutions, it would be costly to implement such a system. Therefore, we have defined our requirements as:

- Being able to compute over encrypted data without the possibility of leaks
- Having private storage that is only accessible from the nodes in the network
- Being able to make calls to APIs

The prospected services of **Enigma Protocol** satisfy our requirements with sMPC implementation. In addition to these services, Colendi has already formed a partnership with Enigma and started collaborating on the development progress. Therefore, we are building our implementation model on Enigma's secure computation environment. Enigma Protocol allows us to split data across nodes and perform computations without leaking information or access entire data sets. The following operations will be performed by sMPC:

- Accessing social media and mobile phone data without exposing the data to the nodes in the network
- Execution of scoring (Machine Learning Algorithms) and updating the score of each user

In addition to sMPC, Enigma also introduces secret contracts that allow us to include sensitive data in safety [6]. More generally, having private data storage means that we can easily separate users' interactions from their IDs *(i.e. Alice may declare that TelcoBob can submit data on her behalf, but Alice and TelcoBob only know this information)*. We will store metadata about the interactions of users in secret contracts as follows:

- User's approvals of data partners
- Data references to files stored in SOS
- SOS Private user keys

First, a user indicates that their data can be fetched from only approved data partners that exist in their allowance mapping *(remembering that in the early stage of the protocol, data partners negotiating with Colendi would be able to push data)*. When someone tries to input new user data, the reference to data will not be appended to that user's references list, unless the user has already approved this.

The latter will consist of an append-only list, with each entry pointing to a reference of an *encrypted/anonymized* JSON blob submitted by data partners. During the scoring process, the algorithm will read this entire list or deliberately selected subset. Considering that user data is stored on SOS as encrypted, in order to execute computations over this data, SOS private keys must be stored in this contract. The user's SOS private key is encrypted with enclave's public key, so it is not possible to decrypt this key out of the bounds of Enigma. In a similar manner, social media tokens may also be stored in this secret contract, but we also have a solution (Section IV-B) that does not require the storage of access tokens.

IV. Data Collection

The Internet's purpose is to ratify knowledge through the accumulation and manipulation of ever-expanding data. Human cognition is losing its personal character. Individuals are turning into data, and data is becoming a center of power [7]. In order to make accurate lending decisions about users, we must collect relevant information from this expanding data. However, relevant information can only be mined from well-targeted data sources. For example, while the amount of daily water a person drinks is not a factor that affects the score, the duration of their phone conversations may have significant relevance. During our research, we have determined three data sources that affect the credit score of a user:

- Mobile phone data
- Social media data
- External data partners

It is crucial to consider mobile phone and social media data, particularly for **unbanked** and **underbanked** people, since we spend more time on our smartphones and share more data than ever on our social media accounts.

A. Mobile Phone Data

Mobile phone data has been studied to make viable claims about users across a wide range of applications, such as inferring friendships [8] and infer socio-economic status in [9]. Computing credit scores from mobile phone activity data are also carried out by academic studies, as in [10]. The industry has also moved towards credit scoring using mobile data (*e.g. Tala has distributed around $300 million loans as in [11]*). In our protocol, mobile phone data is a strong factor to calculate seed score. Even if we don't have any user-related data from partners or social media accounts, we are still able to evaluate credit-worthiness using mobile phone data.

Mobile activity is logged and stored on devices. During registration, users are asked to allow the Colendi App to access these logs. Hence the application starts collecting mobile device data with user consent. Unfortunately, the amount and quality of data vary significantly between Android and iOS; as the former allows significantly more data. iOS does not grant access to call history or the list of installed applications. However Android presents a different case as the application should begin accessing metadata available on the phone that is collecting specific data such as the list of installed applications, recent call history, and cellular data usage. Once this data is collected, it should be bundled by serializing it into a Colendi-standard JSON schema followed by submitting to the network via the Colendi SDK. Call durations or the list of installed applications provide meaningful information about the end user. According to research conducted with 3,760 users, the ones who had installed LinkedIn, Fitbit or Yelp are more likely to have an annual income level above $50,000 per year [12].

B. Social Media Data

Studies in the field of behavioral economics show that people are likely to interact with others who have similar levels of creditworthiness [13]. If it is possible to generate a network of people's daily relations, and interactions (*colleagues, friends and families*), this network would definitively provide reliable data about creditworthiness for each person in the world. In our minds, the best simulation of the

global network defined above can be achieved via social networks such as *Facebook, Twitter, Instagram, etc*. After extensive studies, we propose a novel algorithm which relies on social media relations and activities between stakeholders. The algorithm is detailed in Section V. Collecting social media data in a decentralized system is difficult, as the most popular networks (i.e. Facebook and Twitter) require API consumers to authenticate themselves with credentials (access tokens) that must be kept private to access user data. These private secrets cannot be stored on public ledgers like Ethereum. Yet without mining social media data, we are unable to execute our novel approach to build this network of creditworthiness. We have overcome this problem with a sophisticated approach built on top of Secure Computation Environment, as explicitly shown in Figure 3. The access token will be stored on the user's phone and pushed to SCE whenever the user needs the score to apply for microcredit. Alternatively, these access tokens may be stored in Enigma's secret contracts, so that their scores are calculated even though users would fail to transfer their access tokens.

The steps for fetching data from social media networks are as follows:

1) Request the access token from social media with an authorization code.
2) Encrypt the access token, and call the function in Ethereum that triggers the eval_socialMedia function.
3) Decrypt the access token, then fetch the social media data in SCE using API of corresponding social media platform.
4) Process social media data such as evaluating activity. Finally, store activity scores and user relationships in SCE.

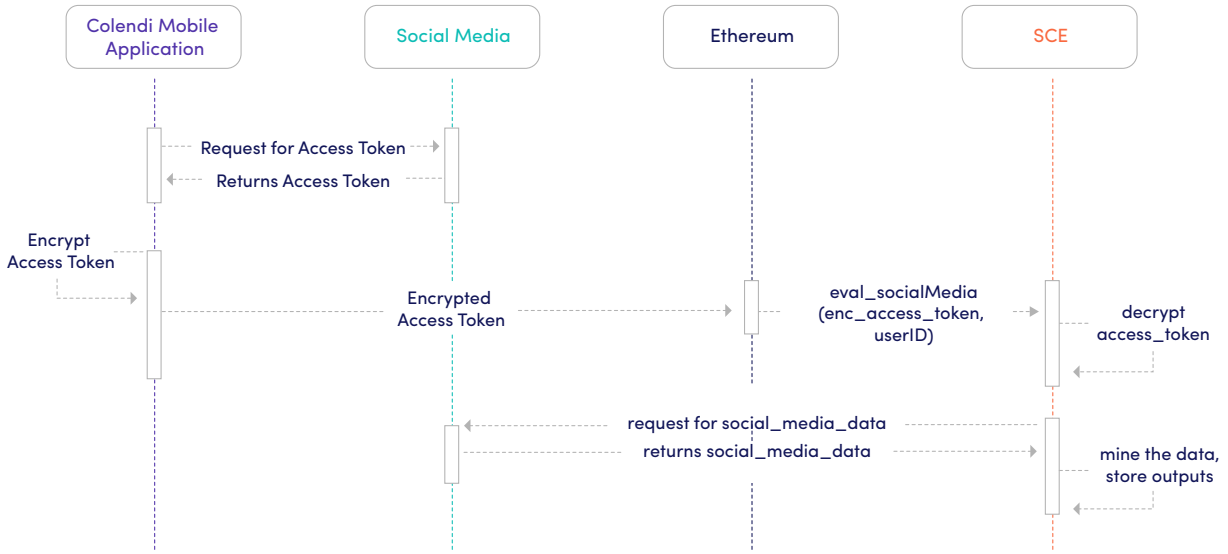

Fig. 3. eval_socialMedia is a function in SCE that fetches social media data by using API of the third party and extracts useful information from returning objects. Outcomes of the extraction are stored in SCE to be used later on score calculations.

C. Data Partners and Data Integrators

Evolution from traditional systems to a decentralized world cannot be realized over a day. During this transitional period, we have to adapt third-party authorities like global/local companies that are commonly considered to be trusted. Their roles are defined as **Data Integrators** and **Data Partners** in Colendi, who

have crucial importance in terms of contributing to the evaluation of creditworthiness. In the initial phase when the participants on the network are yet to be ranked, valuable data will be provided by our data partners to evaluate the seed scores of each user. These participants will consist of entities such as telephone companies, lenders and merchants, and be permitted to submit data about user activities to further enrich the pool of available data for score generation/lending decisions. In order for users to maintain control over their profiles *(and prevent fraudulent data submissions)*, data partners will be required to seek user approval before submitting the data that will be used for scoring evaluation.

As an example of this flow, let us consider Acme Inc., a telco operator. When a user signs up to Acme, they are presented with the option to link their Colendi account. If the user accepts, they will be prompted to scan a QR code that will allow the Colendi App to sign a transaction, stating: *Acme Inc. has permission to submit information about my usage on their network*. This signed statement is then submitted to a sensitive data storage in SCE which records the mapping of Colendi IDs to provider approvals and revocations. From this point on, the data partner may use the Colendi SDK to encrypt and publish data to the network. Note that the approval transaction should use Acme's public key as a verifier, as all data submitted by Acme should be signed with their corresponding private key to ensure input authenticity.

In case the service is terminated, or a disputed data input takes place, users will also have the ability to sign and submit a transaction revoking the corresponding data partner's permission to publish data about them. After users revoked this permission, data partner cannot input a reference to users' data references list. As the reference will not take place in the list, either it will not affect the score calculation of users or data partner will not get a reward for publishing data during an unauthorized period.

V. Score Computation

Colendi Score is the universal credibility measure of each Colendi user. It states the creditworthiness of a user, which provides a viable risk analysis tool for lenders to decide whether to issue microcredit. Each Colendi Score is bound to a corresponding Colendi ID.

The generation of a user's credit score proves to be challenging, due to the requirements of the computational solution we are building: First, such a solution must be powerful enough to process and combine data from a user's stored data. Second, it should be secure and decentralized, meaning that it must be able to run in an untrusted environment without revealing any personal information of the user to third parties. Note that the proposed architecture chooses to side with *user control/privacy* versus *algorithmic intelligence*, as the latter requires access to unencrypted user information. We believe that Colendi primarily adds value by guaranteeing data sovereignty and privacy for the user. For this reason, we propose a solution built on top of SCE for computation and parameter-hiding purposes, while choosing SOS for an end-to-end encrypted storage solution to provide a secure and decentralized protocol.

In addition to these challenges, calculating a credit score without any prior financial data or missing features (i.e. iOS phone) are some other problems to resolve. However, we also propose an algorithm based on the paradigm that people of similar financial repute tend to connect with each other. As already mentioned, this paradigm is the outcome of academic studies on behavioral economics and data science. Additionally, we are conducting tests on a dataset that proves the viability of our novel algorithm.

We are initially required to evaluate a **seed score** for users, for which we need sufficient data to start with. Firstly, the seed scores of these users are evaluated with a Decision Tree algorithm using mobile phone data and input from data partners. The Decision Tree provides an explainable output which can be represented as if/else blocks. So, it can be executed on smart contracts which will allow us to calculate fully-decentralized credit scores in the future.

Having calculated seed score for the subset of Colendi users, we then apply our novel approach to calculating the creditworthiness of users, the ones we don't have prior financial data or have insufficient data features. The scoring algorithm works with users as follows:

1) Generate a network of users (*Figure 4*) as with two directed edges:
   a) Social connectivity
   b) Activities
2) Calculate the score of each person
3) Normalize it to a convergence

Social connectivity strength is the result of social connectivity analyses of a social link between two nodes. It defines social relation strength and is a direct contributor to social credibility. It is a compound of relation scalar strength between two nodes, denoted by $W_{relation_{AB}}$ and dependent bilateral social relation activity, denoted by $W_{activity_{AB}}$.

$$W_{AB} = \beta * (W_{activityAB}) + (1 - \beta) * (W_{relation_{AB}}) \text{ where } \beta \text{ weighting factor.}$$

Having a high score connection in your friend circle does not mean that there is a close relation unless a mutual social interaction is present. This metric also overcomes social network manipulations to achieve higher creditworthiness.

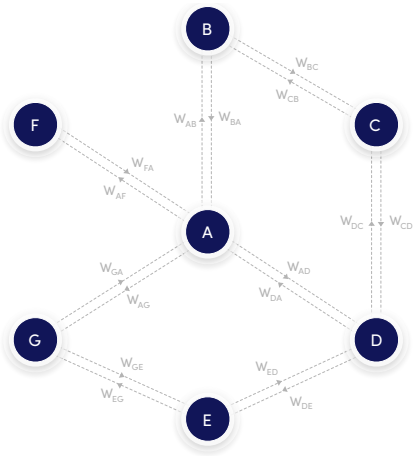

Fig. 4. Connections between nodes

We calculate **ColendiScore** as follows:

$$S_{COLENDI_A} = \alpha * (\frac{(1-d)}{N} + \frac{d * \sum_i (\frac{S_{COLENDI_{A_i}}}{S_{max}} * W_{AA_i})}{\sum_j W_{AA_j}}) \quad (1)$$

where: $S_{COLENDI_A}$ = Colendi Score of user
d = Damping factor
N = Total number of users
α = Normalizing coefficient
$S_{COLENDI_{A_i}}$ = Colendi Score of A's ith neighbor
S_{max} = Maximum Colendi Score of node at graph

Colendi Score for each user is stored on SCE, as with Enigma's secret contract. The contract includes a mapping structure which takes the Colendi ID of the user as a key and the corresponding score as a value. The nodes on the SCE network can only update this contract.

It is mandatory for us to provide up-to-date user scores at all times. Considering that updating all user's data continuously would be costly and unnecessary, we calculate this score in an adaptive way, as described in Algorithm 1:

Algorithm 1 Score Update Algorithm

Input: $X \leftarrow$ initial score update time limit
Output: $\alpha_{user_{i-1}}$ score update time limit for $user_i$

```
∑_{i-1} set(α_{user_{i-1}}) = X
if α_{user_i} > α_{user_{i-1}} then
    if α_{user_i} > Y then
        decrease(α_{user_i})
    end if
else
    if α_{user_i} < Z then
        increase(α_{user_i})
    end if
end if
```

User scores that are kept up-to-date and stored privately can be queried in two ways. First, end-users may always display their scores in their mobile application free of charge. Second, either directly or via merchants, lending platforms can execute score queries to determine whether potential borrowers who applied for microcredit are eligible according to the predefined standards written on a smart contract. In this case, the lending platform/lender is charged with a query fee to gain access to borrower's score.

If appropriately standardized, these scoring contracts can also be created by third parties in a marketplace, allowing lenders to select between Colendi's generic scoring algorithm versus more customized scoring algorithms, which are tailored to their particular needs and customers. Parties may even be incentivized to create/improve on our scoring algorithms by being able to collect the fees as a reward each time their algorithm is used to score a customer.

REFERENCES

[1] A. Demirguc-Kunt, L. Klapper, D. Singer, S. Ansar, and J. Hess, *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. The World Bank, 2018.

[2] P. J. Leach, R. Salz, and M. H. Mealling, "A Universally Unique IDentifier (UUID) URN Namespace," RFC 4122, Jul. 2005. [Online]. Available: https://rfc-editor.org/rfc/rfc4122.txt

[3] D. Reed, M. Sporny, D. Longley, C. Allen, R. Grant, and M. Sabadello, "Decentralized identifiers," 2018. [Online]. Available: https://w3c-ccg.github.io/did-spec/#the-generic-did-scheme

[4] S. Wilkinson, T. Boshevski, J. Brandoff, J. Prestwich, G. Hall, P. Gerbes, P. Hutchins, C. Pollard, and V. Buterin, "Storj a peer-to-peer cloud storage network," 2016.

[5] V. Buterin, "Privacy on the blockchain," 2016. [Online]. Available: https://blog.ethereum.org/2016/01/15/privacy-on-the-blockchain/

[6] G. Zyskind, "Defining secret contracts," 2018. [Online]. Available: https://blog.enigma.co/defining-secret-contracts-f40ddee67ef2

[7] H. A. Kissinger, "How the enlightenment ends," 2018. [Online]. Available: https://www.theatlantic.com/amp/article/559124/

[8] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the national academy of sciences*, vol. 106, no. 36, pp. 15 274–15 278, 2009.

[9] V. Soto, V. Frias-Martinez, J. Virseda, and E. Frias-Martinez, "Prediction of socioeconomic levels using cell phone records," in *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 2011, pp. 377–388.

[10] J. San Pedro, D. Proserpio, and N. Oliver, "Mobiscore: towards universal credit scoring from mobile phone data," in *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 2015, pp. 195–207.

[11] J. Shieber, "With loans of just $10, this startup has built a financial services powerhouse in emerging markets," 2018. [Online]. Available: https://techcrunch.com/2018/04/18/with-loans-of-just-10-this-startup-has-built-a-financial-services-powerhouse-in-emerging-markets

[12] E. Malmi and I. Weber, "You are what apps you use: Demographic prediction based on user's apps." 2016.

[13] Y. Wei, P. Yildirim, C. Van den Bulte, and C. Dellarocas, "Credit scoring with social network data," *Marketing Science*, vol. 35, no. 2, pp. 234–258, 2015.