A Decentralized Infrastructure for Transparent and Competitive Risk Scoring

Traity

***Abstract*—Risk scoring is an opaque, centralized, and narrow industry that leaves a big portion of the population without fair access to financial services. In recent years our credit requirements, ways of living and demographics have changed. We thus need to leverage new forms of structured and unstructured personal data that portray our real behaviours more accurately, which together with artificial intelligence technologies building next-generation risk scoring mechanisms, can help people get access to fairer financial services.**

We propose a decentralized infrastructure for personal scoring that 1) enables data sharing and reward contributions to a whole ecosystem of scoring apps; 2) creates a competitive marketplace for next-generation scoring providers that leverage such data; and, 3) empowers consumers by enabling a plethora of fairer risk scores that will fit better with their lifestyle requirements, their personal reputation and their demographics, while giving them ownership, transparency and control over how and when their data shall be used with full auditability. This system uses cryptography and blockchain technologies to establish the protocol rules, minimize data exposure risks, create the right economic incentives for the ecosystem and act as the API for an ecosystem of applications that can include insurance, credit, sharing economy or fraud.

I. Introduction

Personal risk financial models are outdated: designed in the past century, they were grounded on job stability and previous debt history, encouraging people into a certain lifestyle and behaviours. Using a credit card on behalf of babies to help them build a credit score while they are children [1] is a common practice in the US that frivolously encourages parents into the cycle of debt and creates all sorts of perverse incentives. Similarly, people are advised to get loans in order to improve their credit scores [2], which is contradictory in and of itself.

We no longer consume, buy debt, work or live in the same ways or places anymore, and current financial scores do not take that reality into account. Gig economy freelancers, students or young families are examples of demographics typically left out of the system for not fitting the "right background", despite holding other sources of data that could prove their trustworthiness and could potentially factor into their risk assessments. Migrants suffer simply because traditional credit scoring is limited by national borders and they are forced to prove their creditworthiness from zero [3]. Refugees and people without traditional financial identities have an ever greater challenge. These inefficiencies tend to accumulate due to Matthew Effects [4], and because of this, part of the first world population has no fair access to some services provided by financial systems and a big portion of the emerging world is heavily underbanked.

In the era of Big Data there are better ways to characterize and port our fair trustworthiness. Today, both traditional and new data about us are stored in bureaus that use them to feed outdated models, or are locked in independent silos (e.g. social networks or sharing economy platforms). We aim to create a solution that helps consumers own, control and leverage their data, in line with new regulations like GDPR, and that enables a new generation of risk scores that are more inclusive and better optimised for an ever-changing society, wide and niche use cases, and different demographics. Such algorithms will benefit from more and better data, and at the same time, will be more specific to each use case and can take advantage of innovation in Machine Learning techniques for risk modelling.

Further, despite customer protections laws like Fair Credit Reporting Act (FCRA) [5], credit scores are often used for purposes they were not built for. Our solution also protects customers by giving them control of when their data is used and minimising the effects of data breaches, so that they can be the ones who leverage and benefit from their own data.

II. Background on Society and Trust

Since the beginning of humanity, trust has served as the key enabler for trade and society. Trusting is a social construct that modifies the behaviours of individuals in a society. The same person might not leave a personal computer alone in a busy library in the south of Italy but might do so in Japan. Fukuyama described the differences between high-trust and low-trust societies and their implications in economic development [6]: Low trust societies have weaker business links except with very close people, leading to politics and families within business, reducing competitiveness and productivity. High trust societies like the nordic countries or Japan have more liquid relationships which favor competitiveness and economic development.

Relationships require trusting one another. The ability to trust is based on a shared set of incentives and values, particularly around what is acceptable and what is not. Therefore homogeneous societies tend to be higher-trust societies, and networks of trust suffer a bias towards homophily (similar to problems of segregation [7]). Fortunately, trusting is mostly a problem of asymmetry of information and incentives. Hence trust can be increased by revealing more information about ourselves, having a third party secure the transaction, and introducing homophily in the system (let two people know that they share more similarities than they initially thought) and introducing shared objectives into multi-time games.

Trustworthiness characterises people's ability to be trusted for different use cases and behaviours. Reputation is a measure of one's trustworthiness, which can be both intrinsic (earned, like a doctor who has performed thousands of surgeries) and extrinsic (borrowed, like a new doctor who shows their degree from Harvard). The different signals and performances that people are made aware of about someone's ability and character become that person's reputation. And, much like an established status hierarchy, what one might call a "distribution of reputations" (e.g., from those who have no reputation, poor reputations, decent reputations, and excellent reputations) allows for trust-based exchange [8]. If there's a distribution of different reputation-levels, I can then compare the person with an excellent reputation against others, as reputations emerge to distinguish the peripheral from the best among us [9], enabling the choice of someone whose reputation aligns with my risk appetite. Reputation is mostly a mental accounting metric that people build of other people over time as expectations are met with consistent positive interactions, and jump-started through endorsement by parties who have had previous interactions. This makes reputation highly subjective, volatile, context-specific and difficult to extrapolate or share [10] [11]. Also, good reputation is difficult to store and carry, particularly for migrants who find themselves in a place where nobody knows them, and bad reputation can be easy to hide in open neighbour networks where bad experiences will not be public or shared with the rest of the network. Furthermore, while reputation is an asset we own and regularly leverage to our own benefit and the benefit of those we trust (e.g. job recommendations, introductions to new business partners), it is difficult to "price" its value [12].

In the 1970s, a debt-based form of reputation was standardized by the US banking industry into a single score, the credit score [13]. Creditworthiness is now used not only for credit but also for tenant references, insurance and other services beyond its original purpose. When the Web 2.0 enabled services like eBay and AirBnb, the original idea of personal reputation started to gain traction and its measurement created a certain value that was easier to price [14]. If Lucy is willing to pay $100 for a night stay at a home where the host, Ann, has 5 stars, rather than $90 for a similar home where the host, Bob, has 4 stars, the value Lucy gives to that additional star of reputation for this particular transaction, or the risk premium on reputation, would be at least $10. However, if Lucy would rather stay at Bob's home if he priced the home at $80, Lucy would be valuing the risk premium between $10 and $20. This creates a demand elasticity curve of price over reputation. So it is possible to convert reputation (such as star ratings) into "monetary value", far beyond the traditional scoring systems of creditworthiness based on debt [15].

An antecedent to the concept of reputational risk is the concept of risk itself, which is a mathematical archetype to measure uncertainty, particularly useful for credit and insurance. Insurance is the mechanism for which an insurer covers another party against the impact of an uncertain event that is statistically less likely to happen than the price of the premium the customer is willing to pay for such an event taking place [1]. For insurance services to be profitable, the price premium must be higher than the expected return of each loss, for each customer. In most cases, however, insurance companies do not have enough information of a risk to accurately capture the expected return of each specific individual premium. The default response to this lack of knowledge is to increase premiums and mutualize risk by insuring a sufficiently independent set of customers within pools of similar demographics. Banks face similar challenges with defaults on borrowed money. While some of the risks are asset-based and independent from the consumer (e.g. House insurance suffering an Earthquake) other risks are directly connected to the consumer reputation and actions (e.g. Renter insurance destroying property, Short-term loan defaults, etc).

As technology progresses, insurance companies aim to gather better data that enables the creation of smaller pools, towards an ideal of "Pools of One", where they would have a full understanding of the expected return for each individual premium, for both the asset risk and the consumer reputational risk, leading to better competitive pricing to consumers and more personalized financial services. In order to do so, both insurance companies and banks make use of shared customer databases, or bureaus, such as Equifax, Experian or FICO. Companies also share back their transaction history with bureaus, which build an increasingly better picture of the history of customers. Hence bureaus enjoy high barriers to entry and network effects, and can charge inelastic prices to access their scores. By controlling the data layer, they also monopolise the layer above, the scoring and algorithms built upon such data, leading to increasingly perverse rules and obscure insights (e.g. *"pay monthly between 30%-35% of your credit card balance to increase your score for the next 6 months, and keep using the card"*). This constitutes an industry with high entry barriers that can be better served. In this paper, we propose a solution that aligns incentives for the different players in this scenario by rewarding the holders (banks, insurers, but also mobile phone operators, sharing economy companies, social networks or e-commerce companies) and algorithm-creation companies (which shall be separated into a different layer) while giving power to the owners of data (consumers).

III. Architecture

We propose a decentralized ecosystem of scoring apps which rewards reputation data providers (federations) for their participation. The main idea is that platforms, in order to empower and add new benefits to their users by enabling them to leverage their own data, announce their risk-valuable data [2], and get rewarded whenever risk scores query those data. This enables an open data bureau that reduces the barriers of

[1]Terry Pratchett illustrates the expected returns on risk scoring in *The Color of Magic*: "Well, suppose you have a ship loaded with, say, gold bars. It might run into storms or, or be taken by pirates. You do not want that to happen, so you take out an inn-sewer-ants-polly-sea. I work out the odds against the cargo being lost, based on weather reports and piracy records for the last twenty years, then I add a bit, then you pay me some money based on those odds–".

[2]These could be online banks, insurers, e-commerce companies, or sharing economy websites.

entry to new players, offers a competitive market of scores, and rewards platforms for their contribution.

The architecture of the system is shown in figure 1. Data providers host their own transaction data in a unified, signed format and use the Scoring Smart Contract to enable access to potential scores. This enables data providers to control their data and modify, update, or remove them when desired or required. The Scoring Smart Contract enables discovery of and access to data while charging appropriately when scores retrieve data from data providers. It is the Smart Contract that distributes the payment to the contributing data providers.

A. Identity

Managing personal identity is a challenge as of today. Solving it is not part of the scope of this paper. We have designed an agnostic system that shall operate with current and future solutions, aiming to build on top of the identity protocols that are currently being developed, including blockchain-based approaches like Civic [16], BlockStack [17], Ontology [18], uPort [19] but also Estonia's e-Residency [20], e-mail, OpenID [21] or even Facebook and Google IDs (we believe it is important to be pragmatic and understand that not everyone wants to safe-keep their own keys), giving consumers more choice about how their reputation data is used, and making the data itself more portable. Over time we expect the winners to have higher security levels and added services beyond the current systems (e.g. Social Security Numbers working as very ineffective private keys in the USA).

The minimal set that the scoring ecosystem needs is that users have access to a pair of keys (p_k, s_k) that define their identity. This enables a set of interactions with the protocol which shall give end-users a greater control of what is happening with their data:

- Accepting or rejecting requests to calculate risk scores.
- Accepting or rejecting requests for future reviews about transactions.
- Freely getting scores and data about themselves.

This identity system should be decentralized and cross-border, by linking a blockchain identity to online and legal identities. At the same time, by being blockchain based, it enables the compatibility with a full range of Fintech solutions flourishing in the blockchain ecosystem.

B. Scoring

The scoring layer is the interface to the outside world risk-related services, either traditional or blockchain-based. It is built as an ecosystem that promotes the creation of better risk scores and the sharing of data among them, while protecting the interests of all players.

To enable the next generation of risk scores the ecosystem is designed as a marketplace where competition between different score providers is promoted. Additionally scores can access data available in the shared data layer, that, in a secure way, exposes data that earlier was locked in so-called *data silos*: most of the time the companies generating personal data do not know (or they are not interested in) how to unlock the full potential of its data, while the entities with the expertise on fields such as *Machine Learning* or *Deep Learning* have no access to the data on top of which they could build efficient business models. The result is that the potential benefits for final users of leveraging personal data for other uses are never reached.

All these characteristics together (competition, data accessibility and innovation) result in better risk scores that are more specific to each use case, have more data available and can improve through the use of state of the art technologies.

C. Data

1) Rationale: Risk scores improve the more data they have available. Instead of a centralised entity that holds all the data and therefore the potential to return risk assessments, we propose a solution where each party generating risk data are the ones hosting it. The data is not publicly available, just a pointer of its existence on the blockchain.

This approach fits perfectly with the recent explosion of available data, or the so called *Big Data* era. Exponential improvements in the cost of storing and processing data result in a volume of data that is valuable to the ones generating the data, but could be used by others directly[22] or indirectly[23] for scoring purposes. For example, car sharing companies are generating terabytes of data on driving behaviours that, while certainly valuable to improve the processes of car sharing, could also have a bigger impact in an industry like car insurance. Nevertheless data is currently locked and neither these companies or the final user can benefit from it.

Data providers get a reward for improving the ecosystem each time that they return requested data, but only if it is compliant with a set of rules defined by the protocol and written in smart contracts (in the same way that laws such as the Fair and Accurate Credit Transactions Act set the rules for credit scoring). In this way, we create a protocol that is fair and protects all players. Additionally each party generating data is the one holding it, which enables compliance with data protection and right-to-forget laws, and reduces the impact of data breaches by decentralizing what is today a single point of failure.

2) Ontology: As an ecosystem where data is exchanged and merged between different sources, an ontology is needed to describe those data. An ontology is a formal specification of a shared conceptualization [24]. As such, it can be more simply seen as a common and agreed-upon data structure between the different data providers. These data providers might be using different terms and concepts to refer to similar aspects of transaction data, such as the actors involved, the value exchanged, or the outcome of the transaction. However, these terms can be homogeneized into a common language or ontology, but partners are free to provide data that extends the ontology in any way.

An example of the data being published by data providers is shown in table I. In this case, an imaginary hotel chain (called HotelChain) shares customer transactions using our standard data template. The example shows a consumer who paid the required amount ($300), although they left important damage

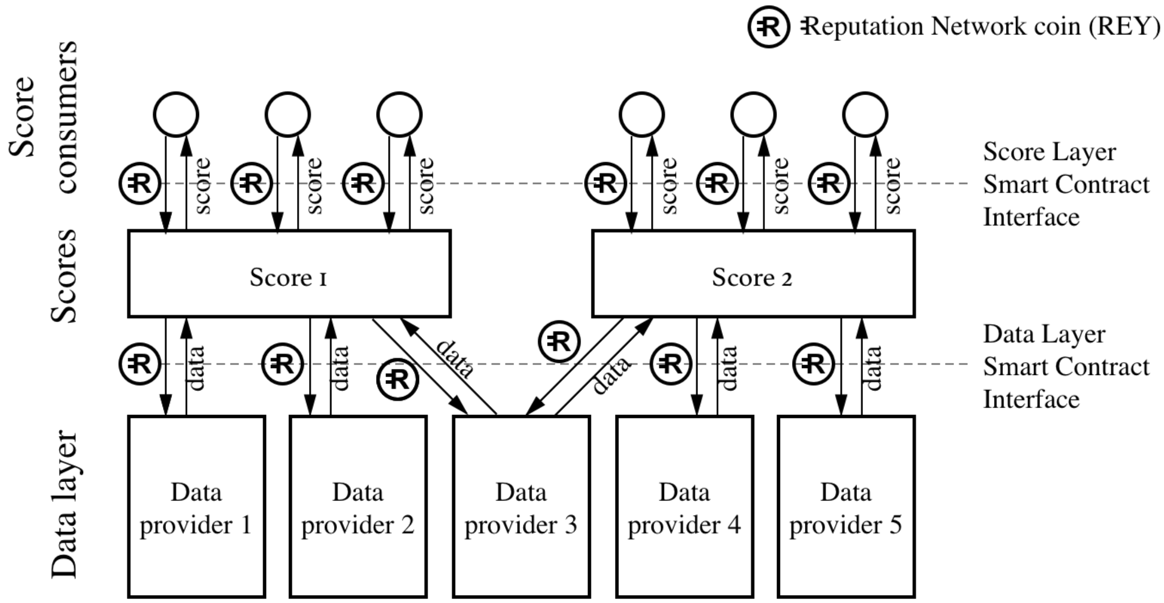

Figure 1. Risk score ecosystem

in the room (which was within the risk that the hotel initially considered).

Transaction ID:	84f6916d8e731c86
Transaction type ID:	hotelchain_stay
Data provider ID:	d4ba55e81a6de5bf
Consumer ID:	c4e08cd8c00ecf5e
Transaction value:	300 USD
Value at risk:	450 USD
Incurred cost:	100 USD
Review to consumer:	★★⯪☆☆
Timestamp:	1517310064

Table I
DATA CHUNK EXAMPLE

	Payment	Risk	Cost	Problem
eCommerce	$90	$90	$90	Credit card fraud
Pay-day loan	$300	$300	$300	Default
Phone insurance	$50	$800	$800	Fake claim of stolen phone
Ride sharing	$15	$15	$15	Did not show up
Hotel	$300	$450	$100	Spoiled room
House sharing	$350	$5000	$1000	Stole TV
Long-term rental	$2500	$2500	$3500	Monthly payment default, plus left damage in the house after eviction
Mortgage	$4500	$4500	$4500	Monthly payment default

Table II
SAMPLE PROBLEMATIC TRANSACTIONS

Also, table II shows different problematic transactions of different kinds using the monetary fields introduced in table I (i.e., transaction value, value at risk, and incurred cost). The model is simple enough to fit very different transactions, from sales to rentals. Most transactions are successful nowadays (the challenge is precisely identifying which ones are not), and thus the incurred cost is 0. However, the table shows problematic cases for illustrative purposes.

As a remark, in some cases, it might be tough to estimate the value at risk, such as in house sharing (platforms might not know the exact value of the house or contents that hosts risk when renting out their houses). Platforms, however, might want to gather such data or provide an estimate in order to provide more valuable transaction data, which scores would like to consume and pay for.

D. Benefits

1) For score clients:

- Any heterogeneous score is welcome in the ecosystem. Some scores might work better for credit, some others might work better for rental insurance, and others for fraud detection.
- A competitive score ecosystem with shared data means a lower barrier to entry to score builders, which would result in broader score offerings. Competition among scores would expectably result in lower prices for the use of scores and better accuracy to increase returns.

2) For score providers:

- Scores can purchase training data sets from the consumers that have made their data available. This lowers the barriers to entry when releasing a score.
- Scores can freely choose which data providers they will use when building, fostering a competitive market for data and rewarding only those data providers that have valuable data. In this case, as there are rewards for data providers, we ensure that (1) big players with highly valuable data will get a higher reward, and (2) newcomers would get no rewards until they publish data. This ensures

a high quality market of data providers without free riders[3].
- By opening transaction data, scores can be built around real world, proven data –e.g. if a user committed fraud or defaulted on credit. This enables the use of Machine Learning techniques over transaction data with a real outcome, unlike pure karma- or points-based mechanisms that are popular on web sites.

3) For data providers:

- A significant competitive advantage for data providers early in the life of the ecosystem will be the ability to differentiate from competing services in their own markets, by being able to claim that they enable their customers to leverage their own data for things that matter to them, as opposed to ring-fencing such data for their sole benefit.
- The ecosystem provides an out-of-the box system that data providers can use to get rewarded and obtain an extra source of income.
- Data providers can benefit from safer interactions on their platforms by means of having consumer reviews that affect scores. This will result in improved behaviours.
- The federated layer of data providers enables providers to control their data in the ways they desire. If needed, they could whitelist/blacklist scores that better match their business vision and ethos. Although the intent of the proposal is that data are available to any score, it is important to note that they retain full control.
- Data providers are free to adjust the reward they get for publishing their transaction data. This means that highly valuable data, in terms of risk prediction capabilities, would be able to charge more to scores. This would privilege valuable transaction data, such as that coming from mortgage payments, over data coming from less consequential transactions, such as gift purchases through e-commerce sites. Thus, data providers will get a direct benefit for data that improves the whole ecosystem. In case the data they give is useless, they will not be used by credit scores and will get nothing in return.

4) For consumers:

- Consumers now have control over when their score and data is queried by proactively giving their permission. The smart contract and the encryption mechanisms ensure scores can only be queried with the user's consent.
- Potential damage due to data breaches is reduced: by decentralizing the data the possible outcomes of a breach are reduced to just parts, and not the whole picture.
- Consumers can request data providers execute any privacy rights from data regulations. For example, if a user demands a data provider delete their data, the data provider simply needs to unpublish it (and remove it from their servers). Pointers to data are stored in the blockchain; the actual data is hosted externally. This solution is compliant with data protection laws like GDPR [25], as the data is hosted by the ones that generate it.
- Data can only be used by risk scores that declare their purpose, and consumers can track when their data is used and by whom.
- Consumers can get an extra source of income if they are willing to share their data for score training purposes.
- Users who are new to a system usually start from scratch with an unfairly low credit score. This is a big problem for migrants who might have a great credit score in their home countries and see themselves forced into large rental deposits, expensive credit cards, down-payments in cash for all services, and waiting several years until they can apply for a mortgage. With this proposed solution, users would be able to port their data and their score, thanks to the unified data format and the open ecosystem. Our architecture makes it easier for a user to leverage valuable online data, be trusted and benefit from it anywhere.

IV. Technology

A. Scoring Smart Contract

We propose a decentralised solution composed of two main elements: a smart contract that establishes and enforces the protocol rules and a network of private channels for data transactions and scaling enhancements.

The *Scoring Smart Contract* defines the rules that govern accessibility, integrity and fairness for all players in the ecosystem. It does this by managing payments between all parties so that only those that play by the rules are rewarded. Rules include an access control layer (that assures previous consent from the end-user for requesting scores and writing feedback) and data integrity and fairness (by storing cryptographic digests of the data at the moment it is announced).

A network of *private channels* manage the peer-to-peer data exchanges, to keep data private and secure. While the minimum viable solution (as explained in section IV-C) uses on-chain payments, the production ready solution will also leverage private state channels to increase speed, improve scalability and reduce fees to the minimum. Then all parties can exchange off-chain *payment receipts* under the same protocol rules which will be checked by the smart contract in the settlement process.

B. Obfuscated identity

User privacy is enforced by the protocol: user permission is required for score requests, their data is stored securely by the ones generating it, scores are calculated without knowledge of the users identity and score requesters do not see more data than the actual score.

Nevertheless, due the public and transparent nature of blockchain technologies, whenever read and write permissions are validated by the smart contract to reward data providers the users public identities get exposed to the public. To guarantee the privacy of the identities behind the permissions being exchanged (e.g. what platforms they use or when they apply to financial services), a scheme of multiple identities is followed.

We define a protocol that leverages *Child Key Derivation* functions[26] to derive a deterministic set of n key pairs

[3] The free rider problem is the scenario where an actor that has not contributed to an ecosystem can get a big benefit out of it

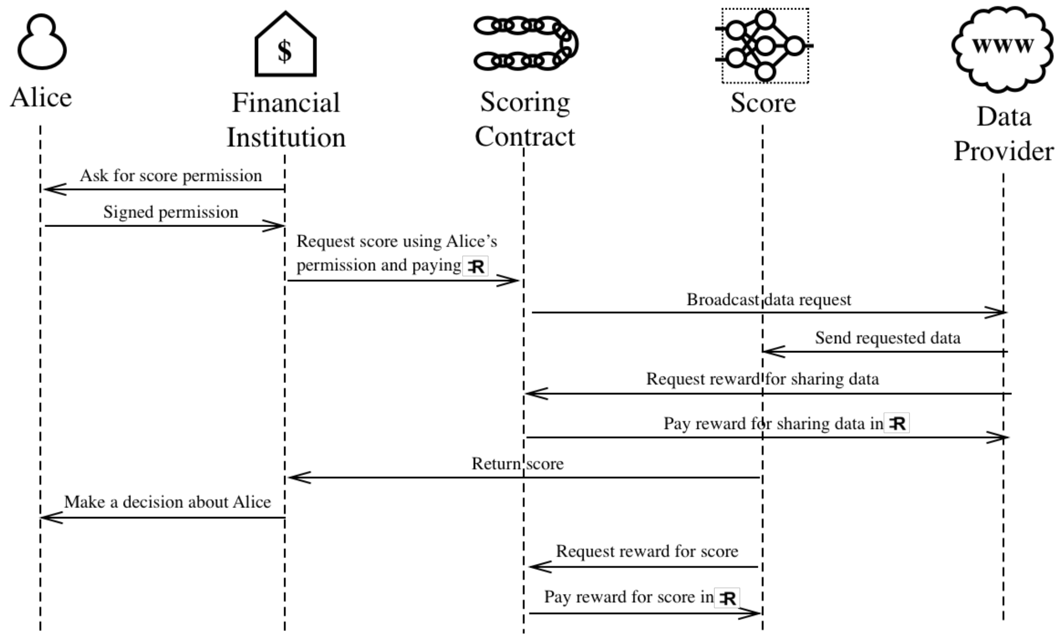

Figure 2. Interaction diagram: requesting a score

$\{(p_{c1}, s_{c1}), (p_{c2}, s_{c2}), ..., (p_{cn}, s_{cn})\}$ from the user identity key pair $(p_k,\ s_k)$. Such child identities are used as follows:

- Data writing permissions are signed with the whole set of child keys, and the identity public key is shared privately with the data provider so it can derive the public key on its own and and prove that the full set of child keys was used.
- Score permissions are signed with a different pair of keys (p_{ci}, s_{ci}) each time.

Thanks to this approach the Scoring Smart Contract will be called with a different p_{ci} for each score request, so only data providers know the real identity (i.e., p_k) of the user.

C. Sample interaction

Let's consider a sample scenario where a financial company wants to query a risk scoring algorithm (figure 2). The participants are: *Alice* as the person that wants to access the financial service, *FinCorp* as the company offering such service, *ScoreInc* as the organization that offers the scores and a number n of data providers $DataInc_n$ that have data about Alice's past behaviours.

1) Requesting a score permission: At some point in the funnel, and before buying the product/service, FinCorp asks Alice for her one-time permission to calculate her risk score. She is faced with the option to decline or accept such one-time permission. If Alice decides to accept the request, she will sign a message using her private key and pass the message to FinCorp.

2) Requesting the score: FinCorp will request the score by calling the smart contract together with the signed permission and the score fee.

3) Answering with data: Each one of the data providers declared by the score provider (that have data tied to a real transaction as explained in section IV-C8) can now return the data using the private channels and use the smart contract to claim their reward.

A data hash has been previously published in blockchain by data providers. The score can then be sure that the data is up to date and that they are the same as other scores might be getting.

4) Returning the score: After the data providers that have data about Alice return their data, ScoreCorp can calculate the score with its proprietary algorithms and send it to FinCorp.

5) Leveraging the score: FinCorp will then use the score to make an informed decision about accepting or rejecting Alice. In case FinCorp only wants to check Alice's risk then the interaction with the scoring ecosystem will be over, but in case she is accepted, she will usually be asked for a permission to write feedback about her in the future (see figure 3).

6) Requesting a data permission: It is in FinCorp's best interest to have the possibility of writing feedback about Alice in the future as both of them will establish a relationship through the service being offered: a loan, an insurance product, etc. It is in this risk-related relationship where FinCorp will observe how Alice behaves which will be useful for related scores in the future. FinCorp has two reasons to offer such data for future risk scores:

- Improving the whole scoring ecosystem by reducing uncertainty and information asymmetry.
- Creating the right incentives for their own users to behave better, as bad behaviours will impact future scores and therefore negatively impact their lives. It has been

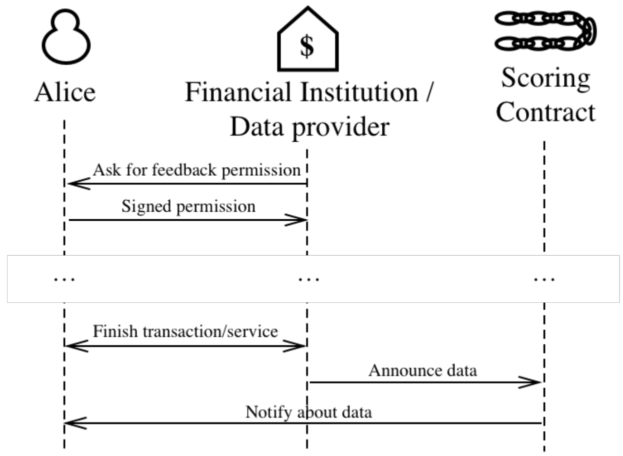

Figure 3. Interaction diagram: adding data to the ecosystem

proven that *"being rated by other players and letting this rating be known are factors that increase cooperation levels"*[27]. This is the basic principle in which reputation systems (like the ones at *Ebay* or *AirBnb*) are built on.

7) Answering with the data permission: If Alice decides to accept the request, she will sign a message using her private key and pass the message to FinCorp.

8) Publishing data: After FinCorp and Alice end their relationship, Fincorp will store the feedback about Alice's behaviours and use the smart contract to announce such data. For that, Fincorp will call the smart contract using the permission and a digest of the data so data integrity can be checked by future score providers.

V. Design decisions and trade-offs

There are several design decisions about the ecosystem that are worthy of discussion.

A. Using blockchain

The score ecosystem is built on smart contracts that run on a blockchain platform, such as Ethereum [28] or NEO [29], the smart-contract-enabled alternative to the notion of blockchain that was introduced by Bitcoin [30]. Alternatively, a fully centralized approach could have been followed. In the latter case, we would assume that all the logic would be running on a server which would be in charge of managing the payments and handling the user permissions.

Obvious advantages would be simplicity in both clients and servers using and serving the system, as well as increased performance. Using blockchain increases complexity, although performance should not be degraded much thanks to using micropayment channels and transporting data off-chain. The main advantage of using blockchain is ensuring the transparency of the rules governing the ecosystem, which increases trust between the different parties. This is a result of decentralization [31]. A couple of malicious behaviours that could be performed by a central authority in a centralized scenario would be:

- Not rewarding all data providers involved in a score request. Depending on the actual implementation, the central authority could even cache the user data, but a simple misbehaviour would be conditionally rewarding data providers from competing organizations.
- Not fulfilling uptime expectations, affecting other parties. Additionally, service interruptions could result in non-executed payments (which could be fixed using cryptocurrency payments, but that would be moving towards a decentralized architecture).
- Not offering trustworthy logs of activity to the other players. The central authority would have privileged information about the scores being returned and the data being exchanged (maybe not about the raw data, but at least visibility of non-public metrics such as the actual requests).
- Privileged access to the score implementations, depending on the actual design and responsibilities of the central server.

B. Using a federated architecture of data providers

A federated architecture of data providers is used, where data providers hold responsibility for handling data and offering them to scores in the platform. Other approaches could have been storing the data in a different system (which could be either distributed or centralized).

One possibility would be encrypting user data and storing them in a publicly available network, or even broadcasting them to subscribed scores. That would avoid requiring to serve user data from data providers, but at the same time would allow scores to cache pieces of data. Similarly, executing the right to be forgotten would be tough in such as system; it would mainly consist of making the data impossible to decrypt by forgetting the required secret, but would leave the data available for attacks, which would pose a privacy threat in forthcoming years if much more powerful machines (especially quantum computers) become available.

In the case that quantum-resistant algorithms are used [32], there would still be companies that would be reluctant to share their user data into any kind of distributed content delivery network, even if data are encrypted. Especially for legal reasons, many companies still prefer to keep private data on their premises, and only share them with those they have explicit agreements with, which makes a federated architecture a better fit.

C. Proposing a score ecosystem

Instead of proposing a score ecosystem, the proposal could have been focused on introducing the next ultimate score. We believe scores are powerful whenever enough quality training data are available, so the focus is on providing an ecosystem for safe data sharing that enables building competitive scores.

In fact, at the moment of writing this paper, there is already a score ecosystem, in the sense that anyone can release a score that uses some data and is useful for some particular application.

However, in a situation where there are several heterogeneous scores around, it would be unlikely that all of the score providers would be reaching agreements with the different data providers to build the most comprehensive score possible. It is also unlikely that all of the scores would be performing a fair treatment of private data, while rewarding data providers fairly. The approach shown in this paper is focussed on introducing a platform that establishes best practices for next-generation scores that compete with one another, and where new sources of data benefit all available scores.

At the same time, as mentioned, we will be releasing one score to bootstrap the ecosystem with readily available functionality. This approach avoids claiming that our score is the fairest or most accurate; it will be up to the market (which involves other forthcoming competing scores and the actual score demand) to decide which score is best.

D. Using a centralized permission system

When it comes to retrieving user data, most current social networks use OAuth [4] or similar protocols to let users control who have access to their data. Such protocols act at data provider level, whilst our ecosystem leans towards a centralized permission system (i.e., one single permission for all data providers) when it comes to reading data[5].

Although following the OAuth approach would provide more granularity and control to users to decide which data providers a score can access, this is precisely counterproductive in a scoring scenario. If using such granularity, bad actors would have control to not give access to data providers where they have done malicious interactions. That would enable bad actors to commit fraud or misbehave in platforms without the scores being able to take into account such bad interactions. That is the reason why the proposed ecosystem requires a single read permission for all data providers.

VI. App ecosystem

Past sections describe a sample scenario where the ecosystem is already up and running, with both data providers and score providers offering their data and algorithms. However on day one there will not be any actors and the project will face a *network effects* challenge: scores will not be useful until there is enough data available, providers will not be rewarded for their efforts until there are enough scores using their data, and consumers will not be there to leverage their risk data.

In order to have a usable starting point, Traity will serve to bootstrap the ecosystem by introducing 1) the first scoring algorithm, 2) the first data provider, and 3) some basic tools to enable exploration for different types of users, including those who might not want to manage their own keys.

[4]https://oauth.net

[5]Still, writing permissions need to be given to each data provider individually for them to publish data about a consumer.

A. Traity

1) Background: Traity is a private start-up with a mission of helping people to trust one another so they can access better opportunities. Traity enables consumers to leverage their reputation from *sharing economy* services, where trust is a key part of every transaction (home-sharing, car-sharing, and other risky peer-to-peer relationships). While sharing economy platforms (such as AirBnb or Uber) are very effective in building trust between their users, they are incentivised to ring-fence their users reputation so that they cannot use it elsewhere. It is in such companies interest to avoid personal data portability so their userbase is effectively locked up: users can not move to a different platform because they will have to build their reputation and trustworthiness from zero. Users reputation, despite being valuable to them (as it can be priced[33][34]), is not "their" property.

With the ambition of changing this limitation, Traity released a cross-platform reputation profile that allows importing reputation data from multiple online sources; Traity users can now extract data from over 25 common online platforms, such as FB, Twitter, or eBay. Traity acts as a data federation in the platform, which can sign the data from those data sources as a proxy of the signature of the original platform. This enables the Reputation Network to operate with several data sources from day 1.

2) The Network of Trust: To give a sense of how companies can build next-generation scores, this section describes Traity's "Network of Trust", patented in 2016[15]; While new sources of data will build an increasingly accurate view of someone's reputation, it is entirely possible that people with a great reputation do not have a significant online presence. Additionally there could be cases where trustworthy users are not represented as low risk: there are infinite dimensions to model personal reputation and by taking a finite representation of a person there will be an intrinsic error present in the outcome (minimising such error will be the objective of every score).

Traity has patented a system [15] to capture information that is not captured by online data: a network of trust relationships where users with a high reputation can altruistically put their reputation at risk to increase the reputation of others, bringing trust relationships from the real world to the online world. This system, that is already working on the Traity website, enables the use of social network analysis (figure 4) to create a *network reputation score* where meaningful trust relationships are captured, rather than just social links like other social networks (*friends* at Facebook or *follows* at Twitter).

Nevertheless the *Network of Trust* brings not only information from the real world not available in existing databases today, but also creates the right accountabilities to improve behaviours in the whole ecosystem: good users altruistically put their reputation at risk to help others, so the recipients of such trust have the incentive of behave well or they will hurt their benefactors. Not only trustworthiness (and thus risk) is better captured but risk is actually lowered for future transactions.

To contextualize this concept, we can say that while we cannot sell our reputations to anyone (they are tied to who we

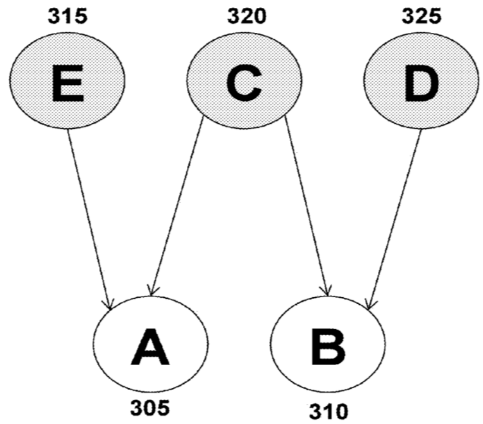

Figure 4. From Network of Trust patent: example of a simple trust network

are), we can stake our reputations with other people (through recommendations). Traity's Network of Trust is another app in the ecosystem that others can leverage.

Thus Traity, as proponent of this ecosystem, shall act as the first scoring provider in the ecosystem, and serve as an example to many competing others that will enable better and fairer services to a plethora of different demographics, minorities, and use cases.

B. Trustbond

Trustbond is the first original data federation in the ecosystem.

Rental deposits are an inefficient form of insurance: People's cash is tied-up in advance for all the damage that could be done in the future. The cost of the rental deposit can keep people from emancipating, or lead to further segregation in cities. Trustbond replaces the rental deposit for a simple insurance premium that frees up tenants' cash, while covering the landlord for the full deposit amount. Trustbond is an example of how reputation data can enable fair and inclusive access to new services and opportunities.

At the end of the tenancy, Trustbond will enable reviews on the Reputation Network. This makes tenants more accountable towards making sure that they leave the property in perfect condition, and will contribute to the profile of those tenants so that they can leverage their own data and prove to any other landlord in the world that they have been good tenants in the past, whether those new landlords want to use Trustbond or not.

C. Bootstrapping

Importantly, while the ecosystem is identity agnostic and handling reputation is not part of Traity's or the Reputation Network's scope, Traity has built tools to help traditional consumers join the ecosystem with friendly user interfaces: a *Login with Traity* button helps encapsulate the whole user experience (from requesting data from users to handling their keys) and a set of *web-based widgets* can be used to easily display and interact with the ecosystem without the need of complex technological integrations.

Traity will open-source some of these data-importing tools and widgets as part of the ecosystem roadmap.

VII. Conclusions

We have proposed a platform that enables a new form of credit and risk-scoring infrastructure. This platform enables:

1) A federated set of data providers to announce their data. These data providers can be banks, insurers, sharing economy companies, e-commerce companies, social networks, mobile operators or other data holders.
2) A separation between data and algorithms layers. These layers are united and owned today by the same bureaus. The separation enables competition, auditability and better incentives for fairer results.
3) Ownership and control by consumers, with a platform that enables permission-based access. While today consumers do not know when their risk data is being accessed, what makes their score and how to modify or improve their scores, the platform enables them to audit and manage their data so that such data helps them achieve their personal and financial objectives.

The applications for this platform include risk scoring for insurance, credit scoring for the banking sector, reputation scoring for sharing economy companies and fraud scoring for e-commerce companies. The different algorithm providers will be able to compete in creating specific algorithms for specific problems or demographics, where they might have additional data to add or a particular expertise that makes them more effective and able to compete, resulting in a better experience and fairer access to services for customers, sustainable rewards for data providers and better prices for scoring clients.

As adoption grows, this proposed platform—through increasingly better machine learning and AI technology teams, more complete and diverse sources of reputational data, and consumers who believe they deserve fairer treatment, security and control, is to deliver a fundamental shift in all trust-based industries—will change how people and institutions trust each other, empowering millions of people to access the opportunities they deserve by leveraging all the reputational history they build and nurture during their lives.

References

[1] The Balance. Making your child an authorized user on your credit card. https://www.thebalance.com/making-your-child-an-authorized-user-on-your-credit-card-960991, 2017.
[2] Experian. Credit tips for all ages. http://www.experian.com/live-credit-smart/tips-for-all-ages.html, 2018.
[3] JP Morgan Chase and Co. Immigrants build credit histories from scratch. https://www.chase.com/news/091815-immigrants-build-credit, 2017.
[4] Matthew S. Bothner, Richard Haynes, Wonjae Lee, and Edward Bishop Smith. When do matthew effects occur? *Journal of Mathematical Sociology*, 34(3), 2010.
[5] Federal Trade Commission. A summary of your rights under the Fair Credit Reporting Act. https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf.
[6] Francis Fukuyama. Trust: The social virtues and the creation of prosperity. *International Journal on World Peace*, 1995.

[7] Nicky Case Vi Hart. Parable of the polygons. http://ncase.me/polygons/, 2017.

[8] Matthew S. Bothner, Toby E. Stuart, and Harrison C. White. Status differentiation and the cohesion of social networks. *Journal of Mathematical Sociology*, 28(4), 2004.

[9] Ronald Burt. *Neighbor Networks*. Oxford University Press, 2009.

[10] Lada A. Adamic, Debra Lauterbach, Chun-Yuen Teng, and Mark S. Ackerman. Rating friends without making enemies. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[11] Rachel Botsman. Who can you trust? In *How technology brought us together and how it could drive us apart*, chapter 5. Penguin Random House, 2017.

[12] Edward L. Glaeser, David I. Laibson, Jos A. Scheinkman, and Christine L. Soutter. Measuring trust*. *The Quarterly Journal of Economics*, 115(3):811–846, 2000.

[13] Sean Trainor. The long, twisted history of your credit score. *Time magazine*, 2015.

[14] Luis Cabral. Reputation on the internet. In *The Oxford Handbook of the Digital Economy*, chapter 13. Oxford University Press, 2012.

[15] C. Herrera-Yagüe, A. Prada, J.I. Fernández-Villamor, and J. Cartagena. Systems and methods for reputation scoring. https://www.google.com.pg/patents/US9363283, 2016. Traity. US Patent 9,363,283.

[16] Civic Technologies. Civic whitepaper. https://tokensale.civic.com/CivicTokenSaleWhitePaper.pdf, 2017.

[17] Muneeb Ali, Jude Nelson, Ryan Shea, and Michael J. Freedman. Blockstack: A global naming and storage system secured by blockchains. In *USENIX Annual Technical Conference*. USENIX, 2016.

[18] Ontology Network. Ontology network: A distributed trust network. https://ont.io/static/wp/Ontology%20Introductory%20White%20Paper.pdf, 2017.

[19] Dr. Christian Lundkvist, Rouven Heck, Joel Torstensson, Zac Mitton, and Michael Sena. uPort: A platform for self-sovereign identity. https://whitepaper.uport.me/uPort_whitepaper_DRAFT20170221.pdf, 2017.

[20] Republic of Estonia. e-Residency. https://e-resident.gov.ee, 2014.

[21] Nat Sakimura, John Bradley, Michael B. Jones, Breno de Medeiros, and Chuck Mortimore. OpenID connect core 1.0. http://openid.net/specs/openid-connect-core-1_0.html, 2014.

[22] Giorgos Zacharia, Alexandros Moukas, and Pattie Maes. Collaborative reputation mechanisms for electronic marketplaces. *Decision support systems*, 29(4):371–388, 2000.

[23] A. Prada. Reputation characterisation using social network analysis. Master's thesis, Universidad Politécnica de Madrid, Traity, June 2016.

[24] Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? In Steffen Staab and Ruder Studer, editors, *Handbook on Ontologies*. Springer, second edition, 2009.

[25] EU Parliament. A summary of articles contained in the General Data Protection Regulation. https://www.eugdpr.org/article-summaries.html, 2017.

[26] Pieter Wuille. Child key derivation. https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki#Extended_keys, 2012.

[27] Riccardo Boero, Giangiacomo Bravo, Marco Castellani, and Flaminio Squazzoni. Reputational cues in repeated trust games. *The Journal of Socio-Economics*, 38(6):871–877, 2009.

[28] Vitalik Buterin. A next-generation smart contract and decentralized application platform. https://github.com/ethereum/wiki/wiki/White-Paper, 2013.

[29] NEO Team. NEO white paper a distributed network for the smart economy. http://docs.neo.org/en-us/, 2014.

[30] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf, 2008.

[31] Vitalik Buterin. The meaning of decentralization. https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274, 2017.

[32] Daniel J. Bernstein. Introduction to post-quantum cryptography. In *Post-Quantum Cryptography*, pages 1–14. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[33] Luis Cabral and Ali Hortacsu. The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78, 2010.

[34] Anindya Ghose, Panagiotis G Ipeirotis, and Arun Sundararajan. Reputation premiums in electronic peer-to-peer markets: analyzing textual feedback and network structure. In *Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*, pages 150–154. ACM, 2005.