
berkly

Berkly

The ***decentralization*** of modern data recollection

Enclo ~
enclo@protonmail.com

01 Abstract

Berkly (BRK) is a compilation of smart contracts (includng an ERC20 token) that live on the Ethereum blockchain with the intended purpose of collecting arbitrary data from different parties/actors without the need to go through a third party to collect and distribute this data. The contemporary method of collecting data (surveys and polls) has the problem that it is subject to human intervention, as surveys are usually run in a centralized manner, leaving the freedom to manipulate data to either the third party providing the necessary tools conducting the survey, or the survey administrator themselves. The idea behind a decentralized platform to conduct surveys has a major drawback however, and that is that there is no central authority to verify the identity of an actor wishing to vote on a poll. This leaves the poll vulnerable to identity attacks where people retain the ability to vote multiple times with no repercussions, effectively destroying the integrity of the data that was being collected.

There is, however, a way to solve this problem - and this is by using the actual token as a form of verification within itself. The idea behind this is that now there is a financial setback for actors wishing to take advantage of the smart contract to manipulate the poll data by having the poll charge the user per vote. This, however, still leaves the poll subject to manipulation for large holders of the token. The fix for this is simple, in combination with the first solution each actor will have the option to add "certificates" to their poll. A certificate will basically be a list, stored on the smart contract, that is controlled by a single address of other addresses it trusts. This opens the door for third parties to conduct identity verifications by themselves on different berkly addresses, and then add them to their list. Then the certificate address can be appended to a poll and the poll will only allow addresses that are trusted by that certificate to participate. It's still decentralized because upon poll creation you can choose any combination of certificates you wish to enforce or allow (or none at all), and anyone can easily setup a certificate. Once a poll is finalized, the results of the poll will be publicly available on the Blockchain for anyone to view and verify, meaning that any party that conducted a poll can prove the integrity of the data that was collected through poll and the public can rest assured that the data that is being claimed wasn't tampered with in the process thanks to the decentralized nature of the project.

Table of contents

02 Introduction

We live in an age where our planet is governed by the exchange and recollection of data. It has become a resource with the same presence in our financial world as money itself. This recollection of data, however, is flawed just like money once was before crypocurrency - in the sense that it is subject to human intervention due to its centralization.

2.1 The problem

The centralization of data collection is plagued by essentially two very big problems. The first is that an intermediary, for example data collection companies, can manipulate survey data. This should seem obvious at a first glance, ofcourse they could manipulate their own data, they are the ones producing it. This however really has no solution with our current systems, someone has to recollect data and that someone can manipulate that data. The second problem (similar to the first) is that the actual survey administrator can claim anything they wish and refer to a data structure which they have no way of proving whether its fradulent or not. It's a system based on trust, and many data collection companies and universities have taken advantage of this trust by rolling out data structures and survey results that they cultivated themselves to emphasize something false to the public. This leaves the public at the mercy of whoever conducted the survey, and even data that is clearly fradulent can't be proven to be so.

The system is clearly flawed, and many of these surveys/polls are very important and their results can have a serious impact on our day to day lives. Take for example a presidental election, how do we, as citizens, have a way to prove that an election wasn't subject to manipulation. Whoever counts the tallies can easily skew data to their liking, and we have no way to hold it against them or prove that they indeed manipulated this very sensitive and fragile data.

This flawed system opens up the possibility for survey/poll systems to be built upon existing blockchains such as Ethereum, however, there is a big problem with systems like these and that is that decentralized data collection structures are simply unfeasible with today's technology due to the fact that there's no "one size fits all" algorithm to correctly verify the identity of an address, leaving the a survey or poll subject to even more manipulation then centralized structures who atleast have the ability to verify the identity of people they are collecting data from. The idea here was to find a solution where the identity verification of different members of a survey isn't based on something like an algorithm and instead on some user-consensus type solution.

2.2 Decentralized verification

The idea behind a decentralized verification system for surveys implies that there is a

surefire way to accuratley prove the identity of a human being 100% of the time, and this is simply for the most part simply impractical. The aim of berkly, however, is to make the act of manipulating decentralized data and verification systems as close to "impossible" as possible. To achieve this, berkly has two main verification types: certificates and stake.

Certificates

The idea behind certificates in hindsight is actually quite simple, but it has very powerful applications on berkly. The idea is that any address on berkly can choose to add another berkly address to their own collection of addresses it trusts. Then at survey/poll creation the administrator of the survey can include any number these addresses **(aka certificates)** and the network will only allow addresses trusted by these certificates to vote. The idea is simple yet powerful because of the fact that a survey administrator can choose to enforce different certificates, meaning that an address has to be trusted by all of those certificates in order to vote. This protects the network against malicious certificates who, for example, can manipulate their list of trusted addresses in order to allow a survey/poll administrator to create fraudulent addresses, get them approved by the fraudulent certificate (which they included in their survey) and be able to vote multiple times. If there are multiple certificates on the survey, you have to convince every certificate administrator to commit fraud - and since the validity of the project is based on the certificates the actual survey used, if unknown or untrustworthy certificates who are likely to commit fraud are chosen, the validity of the data structure collected is instantly nullified by everyone who audits the survey due to the fact that the certificates that a survey used are publicly available for everyone as they are stored ontop of the Ethereum blockchain.

Essentially, it's a system of decentralized trust. Since the survey administrators can also enforce their own certificate in conjunction with other known "trustworthy" certificates, they themselves can check that their own survey isn't being taken advantage of by other certificates by manually verifying their audience with their own certificate (even though this certificate would have no validity to the public auditing the survey in question, if they have other known certificates in the survey it would remain being a "trustworthy" survey). The more certificates that are enforced, the more the public can rest assured that the survey wasn't manipulated.

There are three main ways the network protects against potential identity fraud committed by the certificates: The first one is that a survey can choose to use multiple certificates, whose advantages are described above. The second one, however, is a bit different. A fee in BRK (more on how fees are processed in a bit) is imposed for adding new addresses to this so called "list" a certificate controls. The idea behind having a fee is so the process of adding bulk fraudulent addresses to a list is a lot harder for bad actors since each addition costs them money (berkly). This has several implications, one of them being it can also be used as a simple way to rank these certificates' "trustworthiness", since the more addresses a certificate has - the more BRK that has been invested into that certificate. This fee also serves as a

PoS (Proof of Stake) model of sorts, which is a simple way for different parties to reach consensus on which Berkly certificates and/or addresses are more trustworthy. This very fact serves as a main component for the third way the network protects against certificate fraud, and this is that each address can choose in conjunction to addresses it trusts ones it doesn't trust (or blacklists). Blacklists (as well as whitelists) can also contain the addresses of certificates, meaning that this whitelist/blacklist method also doubles down as a way certificates can perform their own judgement on other certificates on the network. This system discourages fraud by a certificate since by committing fraud they risk getting blacklisted by a collection of certificates who hold more addresses than them (since the amount of addresses held by a certificate corresponds to the amount of BRK that has been invested into it). This makes it so there is a simple way for clients to determine whether a certificate is fraudulent or not, and this is by checking if the amount of addresses held by the certificates blacklisting a certificate outweighs those whitelisting a certificate (in conjunction with the amount of addresses the certificate in question holds). If false, the certificate will show up as fraudulent on the report of all surveys that used it. This, of course, requires voting clients to have implemented a method of checking this (for example, the main berkly website).

Maintaining a certificate can also be treated as a business, as certificates are free to charge a premium to identify addresses and add them to their personal whitelist (aswell as certificates earn a 20% fee of a surveys public voting pool, more on that later). It's for this very reason certificates would be discouraged to perform identity fraud, since they risk being blacklisted by other certificates with more addresses than them, therefore abruptly ending their business since nobody would trust them anymore (excluding the fact that they would require a large amount of capital to commit fraud at all due to the other reasons listed above). Also, every time an address is added to a certificate, the amount of theoretical BRK value the certificate has increased - meaning that this investment would be rendered useless if the certificate got blacklisted by other, more trustworthy, certificates, and it would cost them lots of BRK to create a new un-blacklisted certificate to get back to where they were originally.

Stake

Stake verification serves as the second form the network protects against identity fraud in general, even if all the certificates fail. The idea is that upon voting on a survey/poll an address is required to deposit x amount of berkly into the survey smart contract, which is withheld from the address until the survey finalizes, which at that point it would be returned back to the voting address.

One of the hardest things to falsify is wealth, and throughout berkly it's used to hinder bad actors from damaging the integrity of any survey. With stake verification, it's only used as that, "verification" - meaning that it isn't the same as a fee, since the deposit is returned to the voter. The reason this balance is withheld from the voters until the survey ends is to prevent addresses from voting with a balance, and then moving their tokens to another address and voting again with that same balance. If this stake

is withheld from the start, it becomes increasingly difficult for a bad actor to vote multiple times as they would need to get new tokens altogether for the fraudulent account (as well as getting them approved by the certificates).

The reason stake isn't the only way the network verifies against bad actors is because it is very vulnerable to large holders of the token. The stake and certificate methods are better off complementing each other. Another problem some might see with stake verification is that it's kind of inconvenient having to have berkly on you everytime you wish to participate in a survey, however certificates face the same problem (it is inconvenient to verify yourself various times on different certificates).

Surveys would also be free to choose how much berkly should be staked on the survey to allow the address to vote. As a rule of thumb, the higher this number is the more secure the survey is against attacks (similar to certificates, where more of them increase security).

2.3 Scalibility

One of the biggest problems faced by any decentralized ecosystem (such as berkley's) is the fact that blockchain-based apps have problems scaling. In the case of berkly, one of the biggest problems would be the fact that storage solutions on Ethereum are too expensive to compete with competitor solutions such as data collection companies. One of the main sources of new capital into the berkly ecosystem would be companies wishing to conduct a survey on the network, while adding a "reward pool" which gets distributed evenly amongst the voters once the survey ends. This is fine for big enterprises, however they won't be too keen on paying a huge premium to store their survey on the ethereum blockchain (which at the time of writing this would cost around 1000 dollars just to store 1mb). There is a simple fix for this however.

Hashing

If you have a good think about it, it would be very wasteful to store the entire survey on the blockchain, as its only purpose is for people to verify that what their survey client is giving them is the actual survey that is supposed to be being conducted (like for example what if someone had everyone answer a different survey altogether). What would be the point of storing the entire survey "forever"? All that is needed is to store a form of verifying that the survey presented is the original survey - whom doesn't necessarily need to be stored on the blockchain. If the survey was stored on something else like an off chain centralized server, all that would be needed to be stored on the blockchain is the actual hash of the survey, and anyone would be able to check that the hash from the off chain survey matches the hash stored on the smart contract. Basically, only the hashes would be checked against each other, rather than the actual surveys.

This also solves the problem of storing private information on the blockchain. If only the hash of the survey is stored, nobody would have anything to base what your

answers are about on. Obviously an instruction set would have to be stored on what type of answers are allowed (for example, making sure that a question with answer choices from A -> D doesn't receive an answer choice of E). Essentially only things would end up being stored on the survey, the hash of the survey and a small instruction set on what the actual answer choices that the survey allows are. This also opens up the possibility for conducting "private surveys", in which only a select amount of people know the actual survey and are allowed to vote (whitelisted through something like a certificate).

There is a big problem with this system though, and this is that the actual text of an answer wouldn't be stored on the blockchain, just the id of the answer choice itself (A, B, C, etc). This means that a fraudulent survey client could switch up the answer choices shown to a user to skew data to their liking (or even show a completely false survey to the user entirely), and there would be no way for the public to know if the user voted through a fraudulent client or a legitimate one.

Originally, the thought here was that it could be fixed by having survey clients cryptographically sign every vote that they knew came through their client. This , however, would remain ineffective, as it would give the survey clients a huge amount of power over the network (as well as they would have to be centralized inorder to keep their private key they are using to cryptographically sign the votes private). However, with a little more thought put into it, without network intervention the Berkly network would remain being as responsible for survey clients as networks like Bitcoin are responsible for crypto wallets (in that choosing a fraudulent wallet can get your funds stolen). Basically, the final verdict here is that in order to not interfere with the balance of power within the berkly network, the end user should remain being responsible for who they let sign their votes - not the network itself.

Value proposition

The value proposition is stabilized in a way so that everyone would benefit from the network. This is important for scalability as it provides a strong financial foundation for Berkly.

It all starts with the entity that is wishing to conduct the survey. Usually, this entity could be something like a company that wants to conduct a survey which they can prove wasn't manipulated to the people they wish to present the survey to. These entities would put up a "pool" of BRK on the survey which would then get distributed amongst all voters on the survey when the survey ends (this would primarily be done to attract people to vote on the survey). Voters would want to participate in a survey to get a percentage of the BRK pot offered by the survey administrators, but in doing so they would need to get verified by something like a certificate (in the case they weren't verified by them yet). Certificates would earn money by verifying the identity of addresses and charging a premium to add them to their list of trusted addresses. The berkly network would also grant 20% of the BRK pot offered by the survey administrators to be distributed amongst the certificates that specific survey used,

serving as a sort of a network reward for validating the identities of a collection of people participating in a survey. This reward that the survey administrators would include in their reward pool would be in the form of the actual BRK token, meaning that in the buying of this token they would also increase market liquidity for the token.

To summarize, the financial aspect of Berkly is very scalable as it is designed to provide value to all the different entities that could end up using the network. It presents many different business models to certificates, a scalable solution for companies who wish to prove their survey data, and a way to earn money and emphasize personal opinions for voters.

2.4 Fee structure

As fees are a very important component of the berkly ecosystem (as they are used to keep certificates in check) it was important to find a scalable solution into how they would be handled, as a one size fits all solution would not work here. There are several problems that arise with a scalable fee structure, with the two biggest ones being: The volatility of the Berkly token (meaning that there couldn't be a fixed fee hardcoded on the smart contracts) and how this fee would be distributed amongst the network.

Token volatility, distribution and scalable fees

There's a big chance the BRK token could become very volatile due to various external market factors. This is bad for fees, because it means that there can't be a single fee hardcoded into the smart contracts. If the fee is too cheap, this layer of security the fees propose are rendered useless as it would be very easy to pay these fees and take advantage of the certificate structure specified above. If the fees are too expensive, nobody gets verified and it becomes increasingly more difficult for people to participate in a specific survey.

The solution here had to be something that scales fees to act in-accordance to market movements so that it would remain functioning properly for its intended use - securing and slowing down fraudulent activity on the Berkly network. The idea is simple, yet quite effective. It starts with people being able to stake their coins on the fee smart contract. This staking grants them the ability to do two things, vote on a suitable BRK fee for the upcoming month by voting on anyone's fee proposal and it also gives them a small percentage of the fee pool for the next month (this privilege would only be granted to those who participated in the fee vote). An addresses' vote is magnified by the amount of BRK they have staked on the fee contract, so for example a vote with 100 BRK staked is 10x more relevant than one with 10 BRK staked, same goes with the amount of BRK an address receives (in the sense that you get a percentage of the fee pool that corresponds to the amount of BRK you have stored on the smart contract). If for any reason someone didn't vote but have tokens staked, the amount they would've gotten gets moved to next month's fee pool

The reason that a vote is required to claim your stake fee is because this fee scaling system is one of the main ways the network protects against certificate fraud, as the fees are only payed by certificates (surveys can choose their own fees). It also allows for the redistribution of funds as the fees are split amongst all the voters on the network.

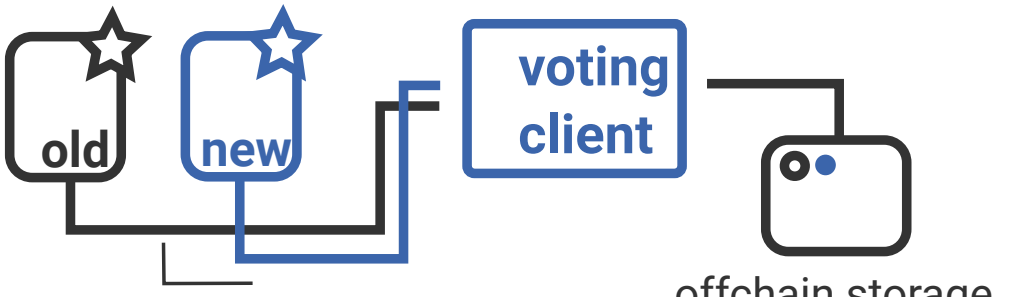
As a rule of thumb, a fee proposal is good if it is cheap for the average person to pay but increasingly expensive for certificates to pay on a larger scale. This would make it so it would be very difficult for certificates to manipulate their list of identified addresses without a large amount of capital (and this would be multiplied by the amount of certificates a survey has). It basically becomes impossible to consistently manipulate trusted addresses on a variety of certificates because of this financial setback, and this is not even including other factors that entice certificates to not commit fraud (listed in 2.2- decentralized verification).

03 Technicalities

Most of the technology on Berkly is simply re-used ideas from other projects put to use to make the project possible, quite frankly, clever implementation of these technologies rather than new ones. That being said, however, some of this stuff could end up being subject to change as the Berkly network continues to be developed. This section will remain brief, as a lot of this stuff was covered in the introduction of this document.

Interoperability and flexibility

Berkly will run on the Ethereum blockchain as a compilation of four smart contracts - not including any smart contracts that could serve as verification certificates in the future for interoperability with other projects. They are, respectively, the fee contract, the survey contract, the ERC20 contract and the certificate contract. The reason these contracts are seperate is so that different parts of the network can continue to work even if in the future a flaw is found in something like the survey contract (as it could be changed in the future by creating a new smart contract, while still consulting the past contracts for the information that was previously stored on the Blockchain). The voting client would then be incharge of interpreting the differences from these contracts in conjunction with the survey stored offchain.

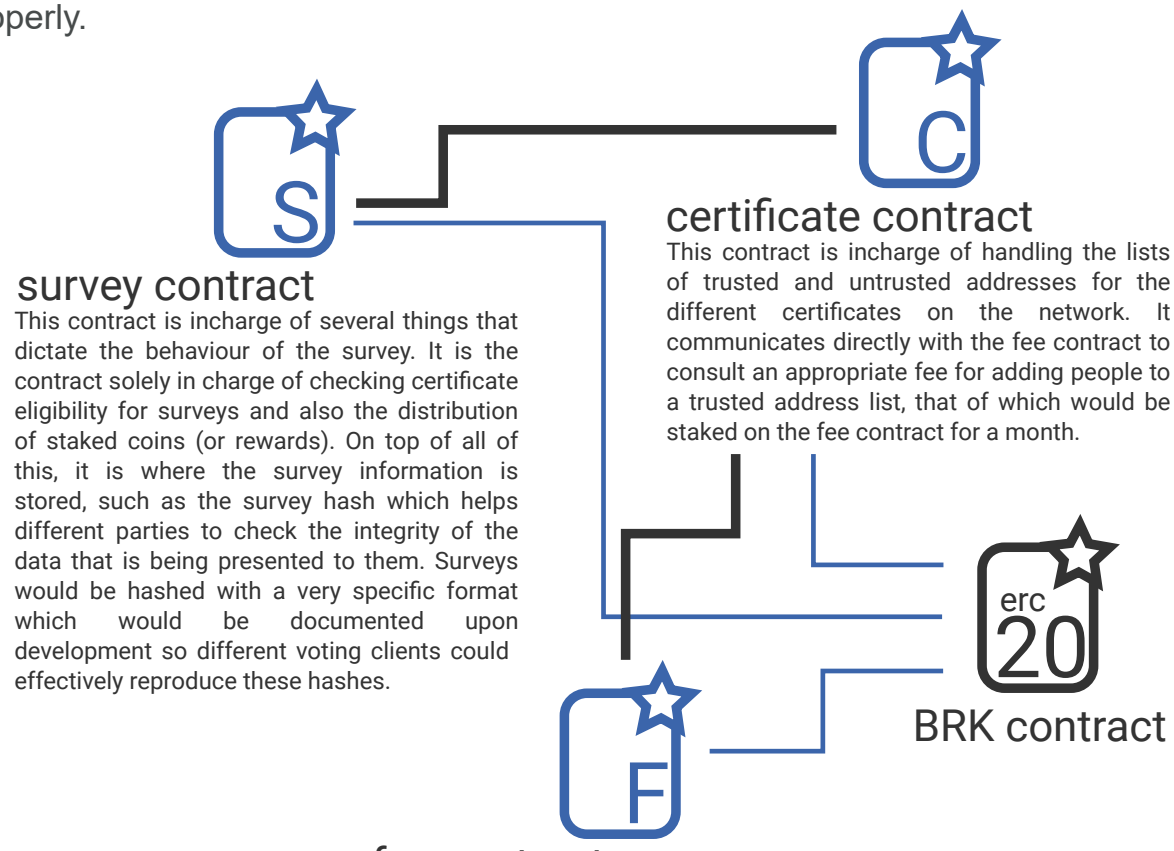


This backwards compatibility with older contracts allows the network to continue to be developed and gives the voting clients freedom to interchange between different versions of the contracts.

This should only serve as a failsafe, however, as the idea behind Berkly is for it to be fully functional from the start - without the need to update the smart contracts. The goal is for development to happen off chain with either something like the voting clients or the actual certificates themselves (who can use innovative technologies to manipulate their list of trusted addresses) and simply have Berkly be a reliable backbone for this activity.

Network anatomy

Several internal transactions occur between the contracts in order for basic network procedures to take place (such as fees). Below is a description of how these contracts interact with each other inorder to allow Berkly to function properly. NOTE: Some of these contracts require external accounts to call special functions every once in a while, such as the liquidation of the fee contract which happens every month (or the finalization of a survey). These functions are public and can be run by anybody as long as it is called after the time period dictated by the certificate - there is no motive for calling these functions other than to keep the network functioning properly.



survey contract

This contract is incharge of several things that dictate the behaviour of the survey. It is the contract solely in charge of checking certificate eligibility for surveys and also the distribution of staked coins (or rewards). On top of all of this, it is where the survey information is stored, such as the survey hash which helps different parties to check the integrity of the data that is being presented to them. Surveys would be hashed with a very specific format which would be documented upon development so different voting clients could effectively reproduce these hashes.

certificate contract

This contract is incharge of handling the lists of trusted and untrusted addresses for the different certificates on the network. It communicates directly with the fee contract to consult an appropriate fee for adding people to a trusted address list, that of which would be staked on the fee contract for a month.

BRK contract

fee contract

This contract is incharge of scaling the Berkly network fee based on community proposals. It is also where all network fees are withheld for a month, and are then distributed to those who have tokens staked on this smart contract and who helped vote on a suitable fee proposal for the upcoming month.

04 Conclusion

Berkly aims to be a way to decentralize the trust on the many surveys conducted everyday worldwide. It gives people a way to cryptographically prove their claims all while giving the public their voice back, an immutable voice that is stored on the blockchain, forever - and can't be manipulated by anyone. The project does have some faults and challenges (which were discussed in this paper), most of them being related to practicality and the centralization that exists within certificates. However, they can all be improved on as the project advances (for example, certificates could be smart contracts themselves, while governed by many different identities through their own consensus layer).

The power to collect and distribute data freely is the biggest value proposition offered by Berkly, and it poses a big threat to many of today's centralized systems that aim to conduct surveys themselves - as it makes it practically impossible for them to lie about survey/poll data.

If there exists truth in data recollection, one can faithfully execute actions based on said data since they know that it can be ***proven to be true.***

berkly

truth in data.