Novo

A protocol for contributing and verifying business data, and a marketplace for connecting data buyers and sellers.

Initial Release: June 1, 2018
Date of last update: September 20, 2018

Abstract

Novo is a decentralized protocol that enables the contribution, verification and sale of structured and unstructured business data via a marketplace. This protocol will support a decentralized application (dApp) built on the Ethereum blockchain.

The NOVO utility token serves as a medium of exchange for marketplace participants, provides incentives for developers and sales teams, and enables decentralized governance.
The token will provide a direct incentive for data providers to sell empirical and other business data through the Novo marketplace where buyers can access data that was previously inaccessible. Additionally, the token will power a protocol to ensure accurate verification of the data. Data verifiers will be rewarded with the NOVO token and data buyers will use the token to buy this verified data through the marketplace.

Cial Dun & Bradstreet (“Cial”), Latin America’s largest cross-regional data provider, will provide data and commit to a minimum of $1 million in purchases to jump start demand. However, the Novo protocol will not be owned or controlled by any single party.

The Novo founding team includes some of the founders of Cial who have experience building a robust network of data buyers and sellers. This expertise will differentiate Novo’s data marketplace as efficient price discovery and data liquidity are assured when buyers and sellers reach critical mass.

Novo strives to be the business data partner for other blockchains, smart contracts, and businesses seeking data on business affiliates and trading partners. Micro, Small & Medium Enterprises (“MSMEs”) will have greater access to credit through verified identity and credit profiles on Novo. Lenders and risk modellers will leverage the data from the Novo blockchain to make more informed decisions. Current business data players will gain access to previously untapped data on MSMEs in developed and emerging markets alike.

Table of Contents

1. Motivation

The likes of Thomson Reuters and Dun & Bradstreet have built sizable companies around large global commercial databases. Analysis of this information helps corporate customers make informed business decisions. Their databases, however, do not include information on MSMEs, especially from developing countries, where data collection is less consistent and verification is costly and oftentimes impossible for non local parties. Those factors lead to poor information about MSMEs in emerging countries. In turn, this has reduced MSME access to credit and limited their ability to grow. The World Bank estimates that $8.1 trillion in credit needs go unmet for MSMEs, primarily due to poor data quality.

The Novo project will address the MSME credit gap by using a decentralized protocol for data sharing and verification. In turn, this will provide individuals in MSMEs around the world an incentive to join the community by contributing their information and verifying the data provided by others. This ostensibly fills the data gap separating MSMEs from the credit they need. The protocol ensures that data providers maintain full control over the privacy of their data. This allows providers to control when and by whom their data can be purchased. These privacy controls ultimately determine how the data is stored and accessed.

Of course, data integrity is of the utmost importance. Once collected, the data must be verified to be accurate. The Novo protocol will assign a group of verifiers to determine the accuracy of the data. Once again, the native NOVO token will be used as an incentive to reward verifiers. An immutable record of the verification will be stored on the Ethereum blockchain. Verified business data will be sold through the marketplace, which will allow for price discovery and yield a return to the data providers and the data verifiers.

The Novo team's expertise at sourcing and verifying data over decades has been employed in the system and verification process design. Several team members and advisors have taken part in successfully building the network of buyers and sellers at Cial Dun & Bradstreet; this expertise will be crucial to ensuring the vibrancy and price discovery of the marketplace.

Instead of each organization having to individually ensure the accuracy of business data it purchases, Novo will leverage economies of scale of purchase, where data can be contributed and verified once by the network, and sold many times.

The marketplace will sell raw data to data buyers, which could be another blockchain, an oracle, a smart contract, a business, or an individual. Additionally, as blockchain infrastructure advances, we plan to offer the option of on-chain data modeling in which raw data can be compiled, analyzed, and packaged into end products directly on the blockchain.

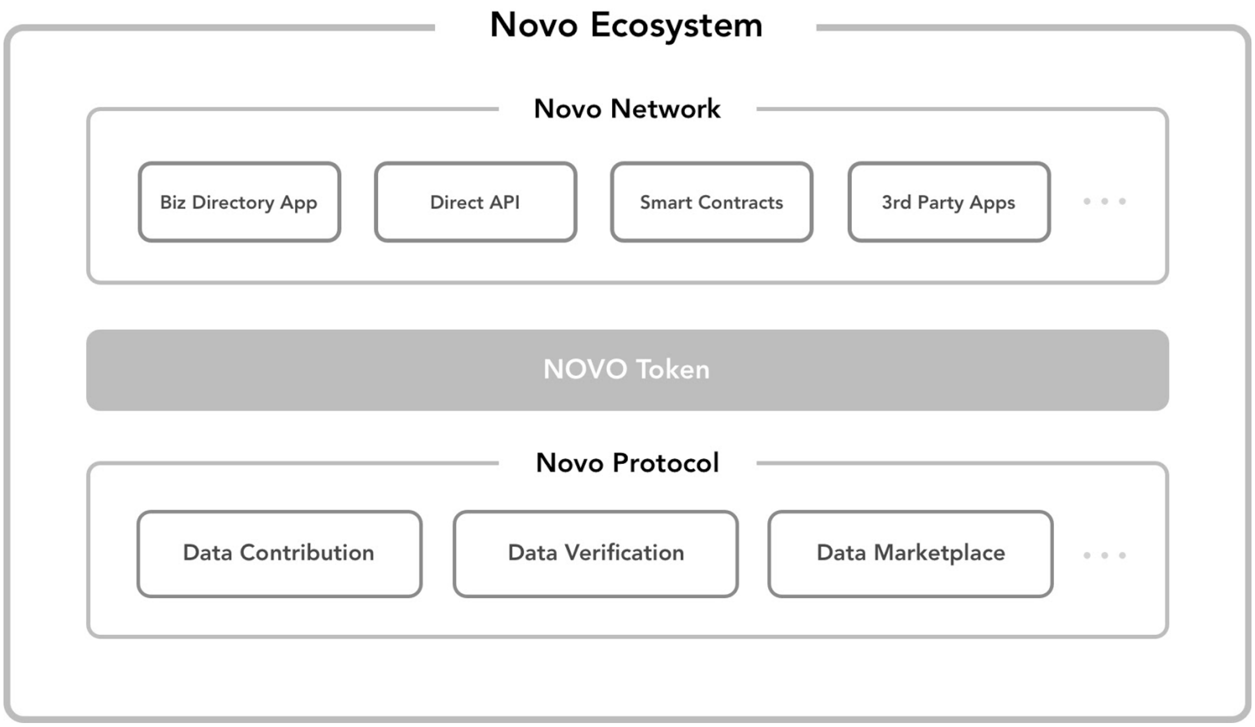
As more business processes become digital, there will be a growing need to reflect real world events on the blockchain. Examples of this are: verifying wallet address ownership, authorizing an employee to transact on the blockchain on behalf of a business, checking a company's risk score before executing a smart contract. Companies depend on credit and flexible payment terms in order to operate, and credit decisions must be based on a company's track record and other indicators. The Novo protocol will serve to reflect real world data on the blockchain.

2. Protocol

2.1. Modular design

The Novo Protocol is comprised of three modular protocols: contribution, verification, and marketplace.

The following diagram outlines the high level design of the Novo Ecosystem:



Novo Ecosystem - Comprises the Novo Protocol, NOVO Token, and all apps and platforms built on top of the infrastructure
Novo Network - The collection of participants that interact with the Novo Protocol by using Novo Tokens, including all apps, APIs, smart contracts and platforms.
Novo Protocol - The base infrastructure of the Novo Ecosystem including modules for the contribution, verification, and sale of data.
NOVO Token - The token required for transacting within the network including the purchase of data.

2.2. Participants

Participants	Function
Data Providers	Submit data about themselves or other organizations
Data Verifiers	Reliably verify data submitted by data providers
Data Buyers	Need accurate informational and transactional data about organizations
App Developers	Develop application layers on top of the blockchain
Protocol Developers	Create smart contracts for the network
Modelers	Create forecasting/modelling tools based on data

▨ Core participants

2.3. Staking

The Novo Protocol utilizes a concept known as “staking” to disincentivize undesirable behavior. A ”stake” is a set number of tokens that must be locked by a participant in order to be eligible to perform an action. These tokens may be lost if that participant is found to be acting in an undesirable manner, as outlined by the protocol.

Specifically, in order for a data provider to submit data to the network, they must first contribute a stake to a stake pool. This stake incentivizes the data provider to act honestly because if the information provided is found to be false or inaccurate, the data provider will lose the stake. Conversely, if the data is found to be accurate and sold, then the data provider will get the stake back in addition to earning NOVO tokens.

Likewise, to become a verifier, each data verifier will contribute stake to a stake pool. If the verifier checks the data and verifies consistently with the majority consensus, and the data is sold, then the data verifier is rewarded in NOVO tokens. But if the data verifier is dishonest and votes opposite of the majority, they can lose their stake.

The details of this model is explained further in Section 4.

2.4. Data Contribution

Let's go through an example.

A small auto parts company in Sao Paulo, Brazil wants to provide information about his company to the Novo protocol. He will submit his stake and provide the data along with the privacy control settings.

The privacy control settings include whether the data is:

Permissioned - The data provider can create a list of data buyers who can purchase the data.
Purchasable - Data is stored after encryption and can only be decrypted and accessed when purchased.
Public - Data that is stored unencrypted, which makes it both searchable and indexable. This can also be considered 'free data'.

The transaction metadata and privacy control settings explained above will be stored in a Novo smart contract. The method of storage will be determined by the privacy control settings. Given the immature state of blockchain technology, data will be stored off-chain with a hash of it stored to the blockchain. However, as blockchain technology improves with advancements in transaction costs, throughput, and privacy, we envision eventually storing all purchasable and public data on the blockchain. With the data securely provided and privacy settings marked as 'Purchasable', the data is now ready for verification.

2.5. Data Verification

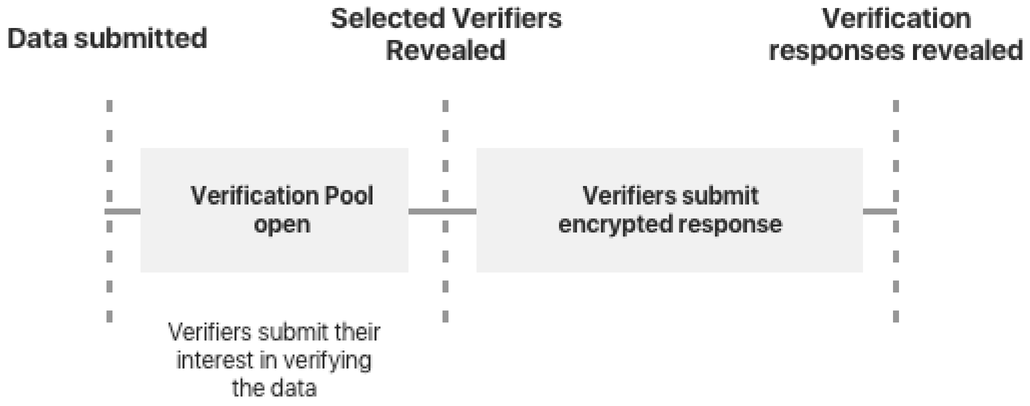
The Novo Protocol utilizes a verification layer between providers and buyers to ensure that all data purchased from the network is accurate. Anyone can be a data verifier so long as they acquire a valid wallet address and stake NOVO tokens, ensuring that they have skin in the game.

To guarantee consistency and defend against malicious actors, the protocol utilizes weighted random selection to select verifiers for each piece of data. Verifiers are selected from a verification pool comprising all those who have staked and opted in to verify a specific piece of data.

A verifier's likelihood of being selected from the verification pool is weighted by the size of their individual stake pool, which represents the total amount that a verifier has locked through staking across the protocol.

The selected verifiers will vote individually on whether the data is true or false. They will not know the votes of the other verifiers so no verifier can corrupt another. Verifiers vote by submitting an encrypted response to the network. Only after all selected verifiers submit their response are the votes tallied. If a configurable threshold, say 51%, is achieved, then the data has been verified and the smart contract changes the state of the data to reflect this.

The following figure shows the process that verification follows:


Data submitted
Selected Verifiers Revealed
Verification responses revealed
Verification Pool open
Verifiers submit encrypted response
Verifiers submit their interest in verifying the data

Let’s return to our example of the Brazilian auto parts supplier. We have a data verifier located in Sao Paulo, Brazil who runs a car dealership. Based on the metadata about the information being submitted by the auto parts company in Sao Paulo, she determines that she has the ability to verify this data. She opts in to the verification pool.

If the data is verified as true, which is the expected result, the data will be available for purchase from the Novo data marketplace. Some public data (including basic contact info) and additional metadata will be available on each company and attribute to ensure that it is searchable by the data buyers.

This is the expected case: data providers provide true data and a group of randomly selected data verifiers votes that the data is true. In a normal distribution, the votes of the majority overwhelms the tails where lie the votes of the malicious verifier (who verifies with malintent), the mistaken verifier (who accidently verifies incorrectly) and the Lazy Verifier (who always verifies true because it is the expected value).

The protocol cannot differentiate the motives of verifiers who vote opposite to the majority but they are all penalized in the same way: complete loss of their stake. To be clear, if the threshold majority of the group of verifiers votes “true”, then all those who voted “false” will lose their stake and all those who voted “true” will earn NOVO tokens when the data is sold.

When the data buyer purchases the data, they are given a set time period to challenge the accuracy of the data along with a stake. Once that time period is over without a challenge, the the purchase amount is distributed to the data provider and data verifiers. The reward designated for data verifiers is distributed through a vesting mechanism detailed in Section 6.

So let’s go back to our example: if greater than 50% of a group of data verifiers votes that the corporate data from the Brazilian auto parts dealer is true, then a hash of the data is added to the blockchain along with metadata about the data including data type and company identifiers. A few days later, the data is purchased by a data buyer who does not challenge the accuracy of the data. The data verifiers who voted “true” are paid with NOVO tokens, and those who voted “false” lose their stake, irrespective of their motive. Also, the Brazilian data provider is now paid with NOVO tokens.

2.6. Example Outcomes

In the following scenarios, the data providers and those data verifiers who vote with the majority are always compensated from the proceeds of the sale when the data is sold. In addition, these parties may earn additional tokens depending on the outcome.

Scenario 1: Data provided is false, verified as false. Data is not available for purchase in the marketplace.
Outcome: Here the data provider loses the amount they staked on that piece of data, which is used to pay the verifiers who voted false. Data verifiers who voted “true” will also lose their stake.

Scenario 2: Data contributed is false, verified as true, and purchased by data buyer who does not challenge the validity of the data.
Outcome: The risk of this outcome is highest for data with small verification pools. Given sufficiently-sized verification pools, this is a bad but unlikely outcome because the majority of a random group of data verifiers are unlikely to vote incorrectly. Additionally, if a data verifier voted False and learns that the majority vote is True, they can challenge the vote within a time period by submitting a stake for the challenge. Once the stake is posted, the data will be open for reverification by a new group of verifiers.

Scenario 3: Data contributed is false, verified true, and purchased by a data buyer who challenges the validity of the data within the time period.
Outcome: When the data buyer challenges, with a stake, the data will be reverified, and if it is found that the data is false, then the data provider is penalized by losing their stake in the Stake

Pool. Those data verifiers who voted True will lose their stake while those who voted “false” will be rewarded with NOVO tokens. The data buyer will receive a refund. If data is re-contributed and verified as true then data buyers will have the opportunity to purchase data again.

Scenario 4: Data provided is true, verified true, purchased by the data buyer who challenges within the time period.
Outcome: When the data buyer challenges with a stake, the data will be reverified. If it is found that the data is true, the verifiers who voted “true” will also be compensated with the stake of the data buyer, while those who voted “false” will be penalized by losing their stake. The reverified data will be offered to the data buyer, who will have lost their stake.

Scenario 5: Data provided is true, verified as false and the data is not available for purchase in the marketplace.
Outcome: Here, the data provider can challenge the verifiers with a stake and the data will be reverified. If it is verified as true the second time, the verifiers who voted “false” in the first and second round will lose their stakes while those verifiers who voted “true” in the first and second round will be compensated when the data is bought. The data provider will receive the penalized stake from the verifiers who incorrectly verified it as false and their second stake will be returned.

2.7. Data Buying

In these five scenarios, the NOVO token is used as an incentive for the participants of the ecosystem to participate with integrity and honesty. Data buyers will also use NOVO tokens to buy the data in raw or modeled form. Here is where the expertise in selling data through CIAL will help determine how to package and combine the data. Data buyers will use the Novo GUI and other applications built on the protocol to search, access, and purchase the data. For purchasable data, only those data buyers who were designated eligible by the data provider may purchase the data. Data buyers may be autonomous, digital entities, other blockchains, smart contracts, oracles, businesses or individuals.

Pricing of the data will be determined by the marketplace and will depend on a series of factors including:

- Time sensitivity of the data: how quickly the data becomes stale
- Velocity of the data: how often the data will be updated
- Size of the data
- Value-added services: provided by the Novo team or by the ecosystem building DApps on top of the Novo protocol

3. Anticipated Challenges

3.1. Data Spam

“Data spammers” post knowingly false or irrelevant information, often in high volume, in order to earn profit at the margins (i.e. via some small percentage of false data “slipping through the cracks” and being incorrectly verified), or in order to cause damage to the network by, for example, overwhelming capacity of the verification pool. Novo looks to limit the scale at which data spammers can operate by requiring providers to contribute a stake prior to posting information. This stake will align incentives of the data providers with the best interests of the ecosystem and dissuade them from spamming the network. The stake is stored in the stake pool as opposed to with the data itself, allowing the stake pool to act as a proxy for reputation.

3.2. Stale Data

Data shelf life is specified by the the smart contract governing each data attribute, including how frequently it should be re-verified. Accordingly, all data must be re-submitted for verification before the validity of the verifications expire. This way, the verified data is always kept fresh.

3.3. Lazy Verifiers

Lazy verifiers assume that data providers, by staking, have little incentive to contribute false information. Because of this, lazy verifiers play the odds and vote “true” without actually verifying. For example, if data providers provide true data 70% of the time, then the true-voting lazy verifier will be paid 70% of the time and lose their stake 30% of the time. This means that if the reward for being correct seven times outweighs the cost of being wrong three times, it is profitable to guess “true.” To prevent this outcome, we must make guessing true unprofitable for verifiers. Novo addresses this through a system called “verification vesting” (outlined in Section 6) where verifiers are only rewarded if they successfully uncover false data, thus eliminating value in guessing true every time.

3.4. Malicious Verifiers

A verifier may have an incentive to purposefully verify data incorrectly (true as false and false as true). Malicious verifiers may mismark data and risk their stake on Novo because they stand to benefit in the real world. For example, a competitor to the Brazilian auto parts supplier might perpetuate false financial statements in order to leave counterparties with the perception that the competitor's business is weaker than it is. If a malicious verifier is mismarking data, they will lose their stake if they are not in consensus with the other verifier(s). However, this cost may not outweigh the real world benefit of causing damage to a competitor. Because of this, staking alone does not prevent malicious verifiers. We solve this attack vector through a combination of staking and random verifier selection from large verification pools.

3.5. Sybil Attacks

Without proper design, the verification system is susceptible to Sybil attacks, where a participant creates multiple synonymous identities to try to take advantage of the system. This creates an issue when it comes to randomly selecting verifiers to verify a piece of data. A data verifier may create multiple identities and apply to verify the same data, increasing the chance of one, or a number, of their identities being selected.

The risks of this type of attack successfully impacting the verification of data are greater within smaller data verification pools, but even large verification pools are susceptible. The problem with relying on the random selection of verifiers to limit the impact of Sybil attacks is that it is impossible to know how many synonymous identities exist within a pool.

There are techniques that make creating synonymous identities more difficult, such as tying real world identities to blockchain identities, but there are still ways to circumvent this mechanism and these techniques require a certain amount of centralized control.

The only true way to prevent Sybil attacks is to design a system in which there is no advantage in creating multiple identities. We accomplish this through a weighted random selection system we call "weighted verifier selection", outlined in Section 5.

3.6. Verification Pool Size

The Novo Protocol design depends on a large population of data verifiers. In the beginning of the network's life, Novo will consider establishing a federation of known and trusted data verifiers. These data verifiers can then invite other data verifiers to the network. Additionally, the governing foundation may provide incentives to early verifiers in select verification pools such as specific geographies or industries.

4. Staking model

Staking incentivizes participants to act in accordance with intended behavior on the ecosystem. The staking mechanisms in the protocol make it unprofitable to act dishonestly within the network by contributing fake data or by verifying data incorrectly.

4.1. Stake Pools

Instead of storing stake separately with each action (contract) that requires it, the Novo Protocol combines all stake for a unique participant. These tokens are locked and will only be released when the participant acts in accordance with the desired behavior in the Novo ecosystem as specified by the smart contract. For example, if a data verifier verifies falsely, they will lose their stake. As a consequence, the amount of a participant's stake will ultimately reflect the value of that participant's reputation. In Section 5, we discuss how this stake pool is used in our Weighted Verifier Selection process to prevent Sybil attacks.

4.2. Stake Pricing

Determining the amount of stake required to incentivize the data providers and the data verifiers is crucial, and can be fine-tuned as the ecosystem grows. The protocol seeks to determine the amount of stake independently for each piece of data, derived, in part, from the market price of the data.

4.3. Stake Duration

Every piece of data in the Novo Marketplace has a configurable duration of validity or shelf-life determined by the data verifier. This shelf-life represents the estimated duration that a verified piece of data can be expected to remain true. At the end of a piece of data's shelf-life, it must be reverified.

Verifiers receive a proportion of revenue for the duration of the data's shelf-life. Of course, if at any point the data is determined to be false during its shelf-life, the verifier is penalized by losing the stake proportional to the remaining shelf-life.

Let's look at an example where a verifier verifies the headquarters address of a business and specifies the shelf-life of this data at 12 months. After 9 months, the business moves headquarters causing the previously verified headquarters address to no longer be accurate.

Because 9 months have passed, the data verifier can withdraw three-fourths (9/12) of their stake. But, if at this time, the data is bought and its accuracy is challenged, the verifier will lose the remaining one-fourth (3/12) of their stake representing the remaining shelf-life of the data.

It is in the best interests of the verifier to accurately estimate the data's shelf-life to balance the revenue received from the sale of the data with the risk of losing their stake if the data is no longer accurate.

Note that a verifier can withdraw their unlocked stake at any time of their choice. Alternatively, they may leave their stake in the Stake Pool in order to enhance their reputation and improve their chances of being selected in the Weighted Verifier Selection process.

5. Weighted Verification

To prevent Sybil attacks (Section 3.5), the protocol will assess the size of an individual's stake pool to determine which verifiers to select, how many verifiers are required and how the token rewards are distributed across verifiers. By using the participant's total stake pool, any benefit from creating multiple identities each with their own stake, is eliminated.

5.1. Weighted Verifier Selection

Verifiers are selected based on a weighted random selection function that weighs a participant's likelihood of being selected based on the total amount in their individual stake pool.

In practice, it works like this: a verifier with a stake pool total of 200 NOVO is twice as likely to be selected than a verifier with a stake pool of 100 NOVO. This means that spreading your participation across multiple identities has no benefit when it come to being selected for a verification. The likelihood of one of the identities being selected is the same whether you have a single identity with a stake pool of 500 NOVO or fifty identities with stake pools of 10 NOVO.

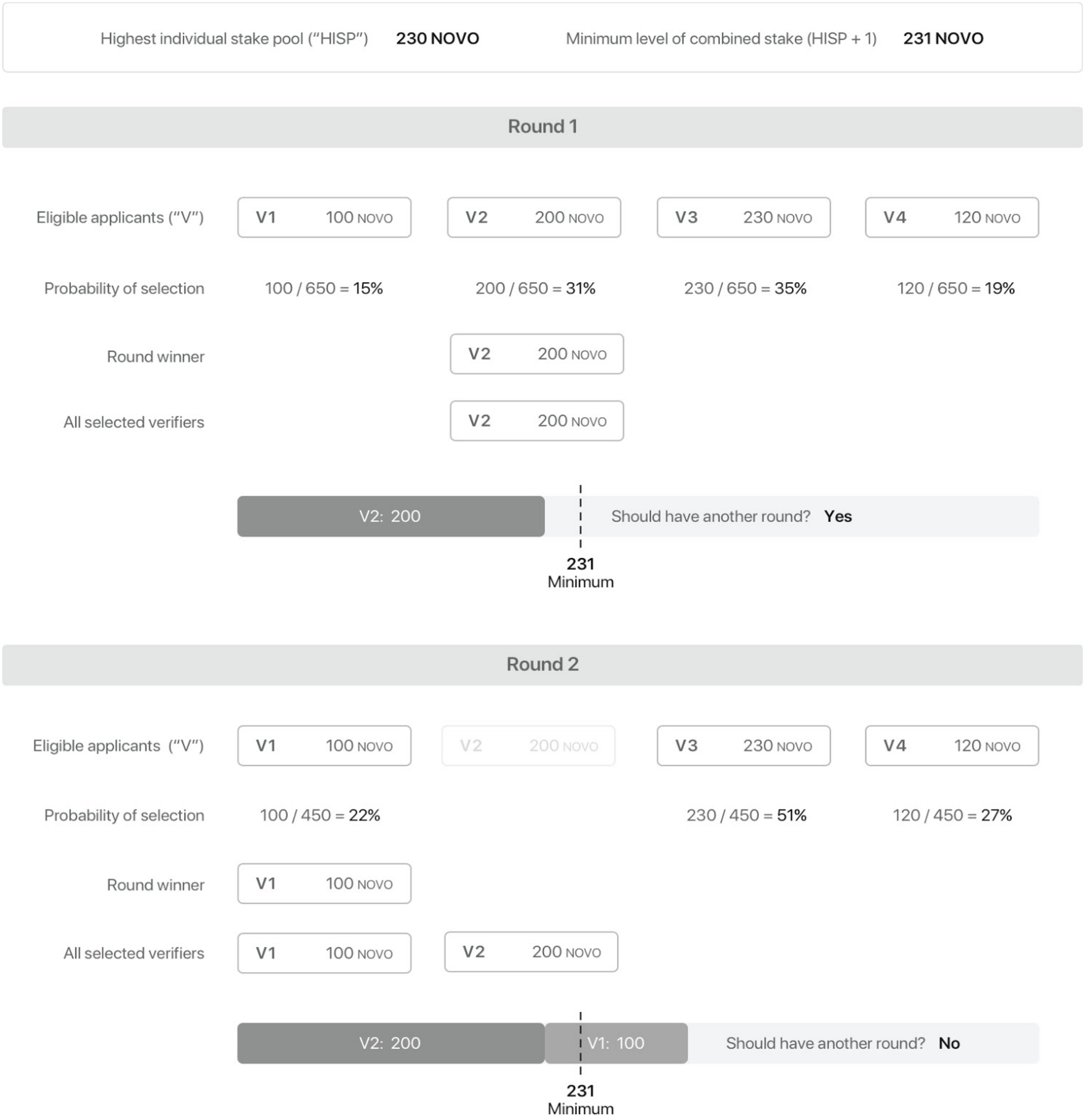
5.2. Stake-based Selection Size

The number of verifiers required for a given piece of data is based not on a defined number, but on a minimum level of combined stake among the selected providers (based on their individual stake pools). This minimum level is determined based on the largest individual stake pool size in the verification pool. Specifically, the minimum is set to one unit higher than the largest individual stake pool. So if the largest stake pool is 500 NOVO and the smallest increment is 1 NOVO, the minimum level of combined stake would be 501 NOVO. This means that if the individual with the highest stake pool is chosen first, one other verifier will need to be selected. This ensures that there is never a scenario where only a single verifier is selected.

Let's look at a scenario where we have a verification pool with verifiers that have stake pool sizes as follows: 100 NOVO, 200 NOVO, 230 NOVO, 120 NOVO. The highest individual stake pool is 230 NOVO, so the minimum level of combined stake would be 231 NOVO. Successive rounds of verifier selections will occur until the minimum level of combined stake is reached.

If in the first round of selection, the individual with a stake pool of 200 NOVO is selected, there would be 31 NOVO left to be filled. If in the second round of selection, the individual with a stake pool of 100 NOVO is selected, there would be 0 NOVO left to be filled. In this scenario, there would not be another round of selections needed.

The graphic below outlines this example.



Highest individual stake pool ("HISP") **230 NOVO** Minimum level of combined stake (HISP + 1) **231 NOVO**

Round 1

Eligible applicants ("V")	V1 100 NOVO	V2 200 NOVO	V3 230 NOVO	V4 120 NOVO
Probability of selection	100 / 650 = **15%**	200 / 650 = **31%**	230 / 650 = **35%**	120 / 650 = **19%**
Round winner		V2 200 NOVO		
All selected verifiers		V2 200 NOVO		

V2: 200

Should have another round? **Yes**

231 Minimum

Round 2

Eligible applicants ("V")	V1 100 NOVO	V2 200 NOVO	V3 230 NOVO	V4 120 NOVO
Probability of selection	100 / 450 = **22%**		230 / 450 = **51%**	120 / 450 = **27%**
Round winner	V1 100 NOVO			
All selected verifiers	V1 100 NOVO	V2 200 NOVO		

V2: 200 V1: 100

Should have another round? **No**

231 Minimum

This may seem less efficient, but it removes the benefit from spreading your stake pool out across multiple identities. The more identities that you spread your stake across, the less each counts towards the minimum level of combined stake required.

5.3 Weighted Reward Distribution

With the stake-based selection size, it is possible to select a large number of verifiers for a single verification. This is good for the certainty of the verification, but bad for verifiers who must share revenue with a greater number of parties. Because of this, if revenue was split evenly among all selected verifiers, it would be in the interest of honest participants looking to maximize revenue to create multiple identities so they had more identities selected. The only way to prevent this is to distribute rewards proportionate to the size of the individual's stake pool compared to the combined size of all selected participant's stake pools.

For example, let's say that 3 verifiers are selected for a verification with stake pools as follows: 150 NOVO, 500 NOVO, 350 NOVO. The combined total of their stake pools is 1000. Therefore, when data is sold, of the proportion that is designated for verifiers, 15% would go to the Verifier with 150 NOVO, 50% to the verifier with 500 NOVO, and 35% to the verifier with 150 NOVO. Without this mechanism, it is in the best interests of verifiers, both dishonest and honest, to have a large number of identities. With these measures in place, we have dramatically reduced the number of identities that are valuable, however it may still be advantageous in certain scenarios to have two identities over one.

Let's look at an example of a verification pool with two participants: Verifier 1 (V1) with a stake pool of 300 NOVO, and Verifier 2 (V2) with a stake pool of 600 NOVO. In this scenario, the minimum level of combined stake required would be 601 NOVO. This means that both participants would need to be selected to meet the minimum. V2 would earn 66% of the revenue, despite filling the majority of the minimum required stake (600/601).

Now let's look at a scenario where V1 still has a stake pool of 300, but V2 has split their stake between two identities with 300 NOVO each. Now, the minimum level of combined stake required would be 301 NOVO. In this scenario, 2 identities will need to be selected to meet the minimum. The best case scenario for V2 is that both of their identities are selected and they earn 100% of the revenue. The worst case scenario for V2 is that only one of their identities is selected and they earn 50% of the revenue. Although, they have the potential to earn less revenue than in the first scenario on any given verification, on average, they will earn more by splitting their stake pool.

The magnitude of this benefit changes depending on the makeup of the specific verification pool and only applies to verifiers with the highest stake in the pool. The important factor that incentivizes one identity over two is that stake can not easily be moved from one identity to another. A user would need to build up two identities over time, using multiple identities even when it is not beneficial to do so.

6. Verification Vesting

6.1. Overview

“Lazy verification”, (detailed in Section 3.3) is when verifiers guess “true” without performing any verification because it is expected that the majority of data contributed will indeed be true. The goal of verification vesting is to reward honest verifiers for true responses while preventing lazy verifying from being profitable.

In this system, verifiers will only earn revenue from their verifications marked as “true” after these verifiers uncover a false piece of data. The discovery of a false piece of data is what “vests” all previously unvested frue verifications.

Until a verification is vested, the verifier will earn unvested tokens every time the data is sold. These unvested tokens will accrue to the verifier, but remain locked until the verification becomes vested. Only when the verifier correctly verifies data as false, are the accrued tokens from previously unvested verifications distributed to the verifier. Once a verification is vested, all future revenue from the sale of the data is considered vested and goes directly to the verifier. Once a verification becomes vested, it remains vested in perpetuity.

However, crucially, if the verifier marks a false piece of data as true, they will lose the right to the revenue from *all* of their unvested verifications. This punishment is purposely harsh because the worst outcome for the integrity of Novo is to have false data verified as true.

Conversely, if the data verifier marks a true piece of data as False, they are penalized by losing their stake without affecting their unvested verifications because the outcome of such a vote is not as severe as verifying false data as True. We want data verifiers to act conservatively when they are marking data as True and not to be penalized for being conservative.

With this system, the lazy verifier who always guesses “true” will never vest, therefore they will never earn any revenue. Those “lazy verifiers” who guess “false” in the hope of triggering vesting, will not be profitable due to the stake penalty associated with guessing “false” incorrectly.

In sum, the goal of verification vesting is to prevent lazy verifiers from being paid when they merely guessed. This technique also helps safeguard against false data being marked as true by extracting a very harsh penalty--the loss of all revenue associated with unvested verifications.

6.2. Verification Vesting Examples

In practice, there are four possible scenarios related to vesting, assuming that the data verifier has a balance of unvested verifications:

Scenario 1 - A verifier responds that data is True, and the data is True
In this scenario, the data verifier votes "true" and the data is indeed true. A portion of revenue from the sale of the data will be reserved for the verifier, but it will remain locked and unvested until the verifier finds a false piece of data.

Verifier's Response	Correct Response	Result
True	True	Unvested

Scenario 2 - A verifier responds that data is false, and the data is false
The data verifier finds a false piece of data and votes False. All previous unvested verifications become vested. Any accrued locked tokens from the sale of the previously unvested verifications will be distributed to the verifier. All future revenue from the sale of the vested data will be distributed directly to the verifier.

Verifier's Response	Correct Response	Result
True	True	~~Unvested~~ → Vested
True	True	~~Unvested~~ → Vested
False	*False*	*Vests all previous unvested verifications*

Scenario 3 - A verifier responds that data is true, but the data is false
The data verifier verifies a false piece of data as True. All unvested locked tokens will be forfeited and all past and future income generated by those verifications would be lost. If this verifier was lazy, and consistently guessed "true" because it is the expected outcome, then they would be caught out this scenario and penalized, making lazy verification an irrational choice.

Verifier's Response	**Correct Response**	**Result**
True	True	~~Unvested~~ → Forfeited
True	True	~~Unvested~~ → Forfeited
True	*False*	*Forfeits all previous unvested verifications (all previous vested verifications remain vested)*

Scenario 4 - A verifier responds that data is false, but the data is true
A data verifier verifies a true piece of data as false. The protocol provides a mechanism for the data provider to appeal a data verifier's decision, and so incentives are designed to entice data verifiers to err on the side of verifying false. This is also precautionary as a way of ensuring that false data is less likely to be verified "true" and impact the integrity of the Novo ecosystem. Here, the penalty for verifying incorrectly is a loss of stake, but unvested True verifications remain unchanged.

Verifier's Response	**Correct Response**	**Result**
True	True	Unvested
True	True	Unvested
False	*True*	*Loss of stake; no effect on unvested verifications*

6.3. Known challenges

We must consider that this system of verification vesting depends on data providers submitting false data now and again. In the instance where all data submitted for verification is true, data verifiers would never discover false data and would never vest their unvested tokens. There must be a sufficient level of false data for verifiers to have an opportunity to vest the locked tokens accrued from their previous correct responses. We anticipate a sufficient amount of false data being submitted for verification due to a number of factors (including human error). However, the protocol can be configured so that there is an intentional contribution of false data which is marked so that even as it makes its way through the Novo ecosystem, it is never added to the marketplace for sale.

7. Conclusion

Many large companies are beginning to understand the potential of blockchain to fundamentally change their business models. They are rethinking existing business processes to incorporate efficiencies made possible by distributed ledger and blockchain technology. This move to distributed ledger-based technology will make more data available. Businesses hold troves of valuable data. The Novo Network will allow them to utilize and monetize that data in a scalable and efficient way.

Companies will be able to expand to regions and data pools that are currently deemed too risky due to informational shortages. Data that is more complete, current and accurate will enable better decisions and accelerate economic growth.

Novo's data marketplace and ecosystem will play a critical role in driving global business forward, both on blockchain and off.

For more information:

Visit www.novoprotocol.com or

Email greetings@novoprotocol.com

References

https://www.forbes.com/sites/forbespr/2017/05/31/poor-quality-data-imposes-costs-and-risks-on-businesses-says-new-forbes-insights-report/#6a0a7b1e452b

https://mastercardcenter.org/insights/imf-mit-study-shows-financial-inclusion-drives-economic-growth/

http://usblogs.pwc.com/emerging-technology/blockchain-and-smart-contract-automation-an-introduction-and-forecast

https://www.civic.com

https://www.uport.me

http://www.cib.db.com/insights-and-initiatives/flow/Know_your_customer_The%20complexities_explored.htm#gsc.tab=0

https://www.thomsonreuters.com/en/press-releases/2016/may/thomson-reuters-2016-know-your-customer-surveys.html

https://www.goodfinancialcents.com/peer-to-peer-lending/

https://www.reuters.com/article/us-usa-banks-automation-insight/automated-lenders-threaten-to-eat-banks-lunch-idUSKBN0OR0BC20150611

https://blockgeeks.com/guides/what-is-blockchain-technology/

https://www.coindesk.com/zk-snarks-everywhere-ethereum-privacy-tech-hits-tipping-point/

https://aragon.one/

https://www.imf.org/external/np/seminars/eng/2014/trade/pdf/ahn.pdf

https://www.insper.edu.br/wp-content/uploads/2012/11/2012_wpe277.pdf

http://www.vicentecunat.com/TC%20Chapter_OUP.pdf

https://www.wto.org/english/res_e/reser_e/ersd_sem_pres_8032017_e.pdf

https://www.imf.org/en/Publications/CR/Issues/2017/03/10/Cluster-Report-Trade-Integration-in-Latin-America-and-the-Caribbean-44735

https://www.imf.org/~/media/Files/Publications/CR/2017/cr1766-ap-5.ashx