July 2018

Cooperative Data Handling in the XYO Network

Erik Saberski[1,2], Arie Trouw[1], Carter Harrison[1]

In order for the XYO Network to operate at maximum efficiency, data must be distributed amongst many platform users and components must work together with little redundancy in their tasks. To achieve this, an incentivization plan must be made to encourage platform users to share data and work together to answer queries, while disincentivizing non-conforming behavior. In this paper, we discuss how the XYO Network encourages our platform users to share data and work together to minimize redundancies in tasks, while reprimanding those who do not increase system optimality.

Data collected from Sentinels is stored on a network of archivists; computers that store, share, and validate data that is used by Diviners to answer queries prompted by end users. Archivists are rewarded for providing good data to Diviners, flagging incorrect data in the Archivist Network, and sharing data with other Archivists. This is all done through utilization of a staking protocol which we term "Archivist Stake."

Archivist Stake (AS) is an amount of XYO that a given Archivist is willing to wage on the fact that they are only holding good and valid data. This stake is in the form of a vest that an Archivist can either contribute to or remove from over time. However, it takes 2 weeks to either add or remove XYO to a given Archivist's AS, where the XYO is either added or removed at a regular and linear interval over the two weeks. This delay prevents Archivists from quickly adding XYO to their AS to perform non-ideal behavior in our blockchain, or alternatively quickly removing their stake when bad data is found in their data store. However, rewards and punishments for network activity can be instantly added to or taken away from an Archivist's AS.

Archivist Data Validation

Part of an Archivist's job is to constantly access other Archivists' data stores to validate and take their data. There are many algorithms to verify data in our network, and as the XYO Network develops there will likely be many new algorithms to verify if data is correct. One, for example, is to check the hashes on each piece of data. When two Sentinels interact, they share the hash of their previous interaction. If an Archivist wants to validate the data with this algorithm, they compare the hash of the data previous to it, re-hash the previous block of data and compares the hashes. If they are the same, the data is considered valid. If the two do not share the same hash, however, the Archivist that is validating (Archivist A) flags the piece of data in the other Archivist's store (Archivist B).

Once data is flagged, 1 hour is given for the Archivist Network to reach a consensus on if the data is valid or not. This is done by the flag being announced on the main blockchain so that every Archivist can see. Next, each Archivist in the network can either vote that A is correct (data is invalid), or B is correct (data is valid) by sending their signed vote to the corresponding Archivist. Since the votes are signed, they cannot be forged, and because they are sent to an individual Archivist, the vote is blind. Votes are weighted by AS, so an Archivist with 10 AS will have ten times the voting power of an Archivist with 1 AS. After 1 hour, Archivists A and B publicly share how many stake-weighted votes they each received, and a winner is determined. At this point, stakes are redistributed with all of the Archivists that voted for the *incorrect* Archivist losing a portion of their stake, and these stakes being redistributed to the Archivists that voted for the *correct* Archivist, weighted by AS. Gaining or losing AS in this manner happens instantly. Thus, it is greatly in an Archivist's favor to only hold correct data, or else they will likely quickly lose their stake.

An Archivist, however, can only be punished for keeping data in his published store. The important distinction between an Archivist's *published store* and *unpublished store* is that data in the published store has been signed by the Archivist. An Archivist can hold as much invalid, incorrect, and/or corrupt data as they want. However, they are only reprimanded (or rewarded) for owning it if they put their signature on it. Thus, Archivists can take as much time as they need to validate data in their unpublished store before moving it to their published store. If they find bad data in their unpublished store, they will likely backtrack to the Archivist that they got the data from to flag it in the other Archivist's store before deleting it from their drive. With this validation system, it can be assumed that Archivists will always hold good data. However, Archivists with higher AS are more likely to hold good data because they are more likely to thoroughly validate their data before publishing it, whereas an Archivist with low AS might immediately publish data too quickly to increase their AS. Alternatively, an Archivist might have a low AS because they have recently been found to have bad data and lost a lot of their stake.

Archivist Share Incentive

As Archivists are accessing each other's data stores and validating data, they are also collecting any data they want to store. However, Archivists may not want to share data as it could create unwanted competition over that piece of data. For example, if Archivist A is the only Archivist that has a specific piece of data, then when that piece of data is needed to answer a query, he is guaranteed to get the entire reward for having that data. However, if they share it with one other Archivists, one might assume that there becomes a 50-50 chance that he will get that reward. This is not true, however, because if twice as many Archivists hold the piece of data, it is twice as likely to be found by a Diviner, and the reward for having that piece

[1] XYO Network, 1133 Columbia St. San Diego, CA 92101
[2] Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, CA 92093

of data is shared across all Archivists that hold it, weighted by AS. Note that data utilization is also proportional to AS, so these numbers are specifically for the case where an Archivist shares with another of equal AS. However, this still works with discrepancies between the two Archivists' AS. Thus, if an Archivist shares data with another Archivist, the likelihood that the piece of data will be used increases by a factor of *x*, but the reward that Archivist receives decreases by a factor of *x* when that data is used. With this, there is no incentive or disincentive to share data. Thus, we introduce the Archivist Share Reward.

The Archivist Share Reward (SR) are XYO tokens that are distributed to sub-networks of Archivists that a block of utilized data is being held by. The SR is first divided amongst the utilized data, weighted by the sum of the stakes of the Archivists holding that block of data. It is then divided further to each individual Archivist holding the data depending on how much that Archivist shared that data. This is shown in figure 1. The reward a specific Archivist receives (r_s) for a given piece of utilized data is given by:

$$\boldsymbol{r_s = \omega R_s \frac{U\, S_d\, \Phi_A}{\sum U \sum S_d \sum \Phi_A}} \quad (1)$$

$$\boldsymbol{\Phi_A = \sum S_{t_s} + \sum S_s} \quad (2)$$

where R_s is the total share reward, $\boldsymbol{\omega}$ is a normalization factor, U is the number of times that piece of data was used that day, S_d is the sum of all AS of the Archivists holding the given piece of data, S_{t_s} is the stake a given Archivist has at the time they share the data, and S_s is the stake of the Archivist with whom an Archivist shares data with. The summation represents the sum of the products of the values in the network for that given distribution period. With these equations, every time an Archivist shares data, they only increase their r_s, regardless of how much stake the Archivist they share with has. However, this does create competition to share with as many Archivists as fast as one can because once other Archivists start sharing the data more, the individual's proportion of the share reward begins to decrease. Thus, sharing data throughout the entire Archivist network is strongly incentivized.

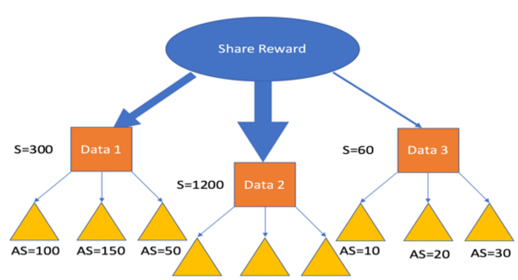

Figure 1. A visualization of the Archivist Share Reward being divided amongst Archivists. Data with greater sum of AS of Archivists holding it get a larger portion of the total reward, as shown by larger arrows.

Archivists also maintain a chorological Proof of Origin chain describing where they get data from. If Archivist A gets a piece of data from Archivist (or bridge) B, Archivist A and B record a hash of this data along with signatures from both A and B, the ID of the data provider (Archivist B in this case), and the hash of the previous block in their origin chain. For each piece of data in an Archivist's published store, it should be able to be traced back to a given bridge through a series of Archivists' Proof of Origin chains. If any data in an Archivist's published store cannot be traced back to a bridge, it is at high risk of being flagged as bad data.

The sub-network of Archivists holding a given piece of data is determined by following the chain of signatures on the piece of data, as well as the Proof of Origin chains. By default, a piece of data will have the signatures of the last 5 Archivists that it was shared with (if it has been shared that many times). This allows the data to be traced even if an Archivist in the chain of data transfers is not online. Archivists also have a record of whom they have shared specific data within their Proof of Origin chain. This allows a Diviner to follow the flow of data both forward and backward to find most of, if not the entire sub-network of Archivists that have the specific data.

Diviner to Archivist Interactions

It is the Diviners' job to read queries, find relevant data in the Archivist Network, and utilize the data to produce an answer. Naturally, Diviners would separately respond to a given query, and compete to provide the correct answer, where the Diviner with the greatest hash-rate is most likely to win. This would be similar to the bitcoin mining protocol where there is only one reward recipient per block and statistically it will most often be the one with the greatest hash-rate. However, this strategy does not improve the accuracy of query answers as our network grows; it only increases unnecessary competition. Instead of having Diviners look through the same Archivists to compete to answer a given query first (i.e. all doing the exact same work), we incentivize them to look into separate Archivists and work together to reach consensus on a query answer. With this strategy, the more our network grows, the more data will be utilized to answer a query, and the more accurate answers will be. Also, more Diviners are rewarded for their contribution -- similar how bitcoin miners are rewarded in a pool.

Diviners can assume that data on an Archivist is correct given the Archivist data validation protocol. However, it can also be assumed that Archivists with higher AS are more likely to hold correct data since they are waging more on that fact. An Archivist with less AS either recently was found to have bad data or is risking a lot because they have little to lose and do not check their own data before publishing it. If a Diviner has a higher hash rate, they will go through the Archivist's data faster, thus it is more likely that relevant data will be found before the query is closed. Since the Archivists with higher AS are waging more and are more likely to have good data, they should be searched through by the Diviners with the greatest hash rate. To incentivize this balance, Archivists set a threshold for Proof of Work (PoW) that a Diviner must do in order to access their data. The Diviners hash the query with a seed to try to get the number of leading zeros requested by their target Archivist. Simultaneously, Archivists have the option to lower (or raise) their difficulty over time to increase their chances of being connected with a Diviner.

The Archivist needs to be given both the seed and the query from the Diviner to verify the PoW. With this information, the Archivist knows which query Diviners are trying to access them for. To reduce redundancies, if an Archivist is found to allow multiple Diviners access them to answer the same query, the Archivist is not rewarded.

When a Diviner and Archivist connect, they first have a signed interaction so that both parties have confirmation that this Diviner is accessing this Archivist. They both store this in their respective origin chain in a similar manner as interactions between two Archivists. Once a Diviner finishes searching an Archivist for data about a given query, they share the data with all of the other Diviners working on the same query along with their signed interaction showing the permission from the Archivist to access their data to work on this query. This gives each participating Diviner identical sets of only relevant data. From this set of relevant data, each Diviner computes an answer and shares this answer with the rest of Diviners. Once 51% of Diviners working on this query agree on an answer, it is used as the solution for the query.

With this protocol, Diviners with low hash rates (cell phones, old computers, etc.) can still contribute in a meaningful way to the network and be rewarded for it. The weaker Diviners will connect with Archivists with lower AS, which still may have extremely valuable data for a given query. Also, as the network grows to have more Archivists and more Diviners, the answer accuracy and speed will only increase.