Oracles for Number Values

William George
Research Lead @ Kleros

william@kleros.io
@williamhwgeorge

HOSTED BY THOMSON REUTERS INCUBATOR

European Commission | Winner of the Blockchains for Social Good Prize from the European Union's Horizon 2020 research and innovation programme

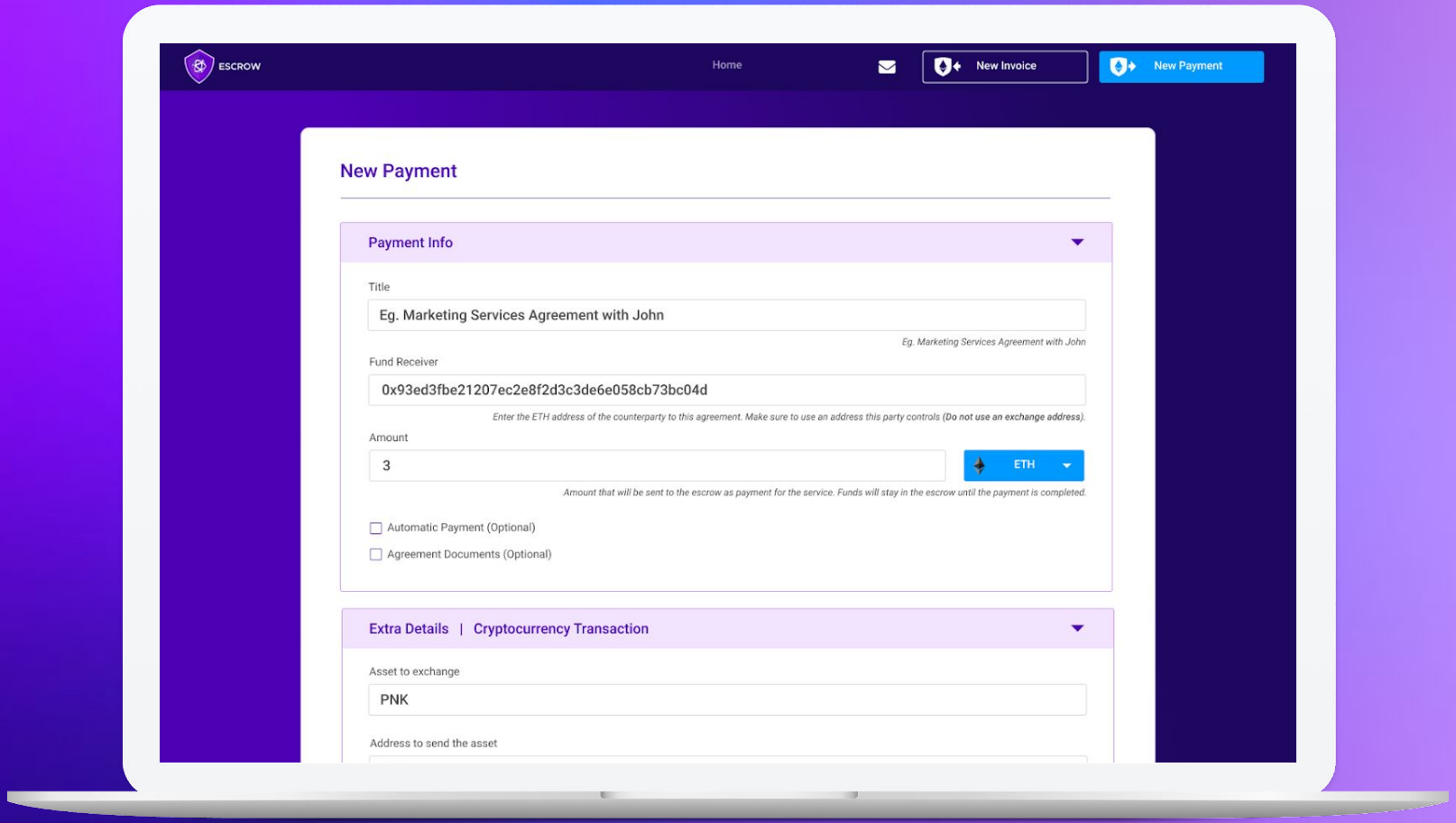
bpifrance

BLOCKCHAINGOV

Motivation

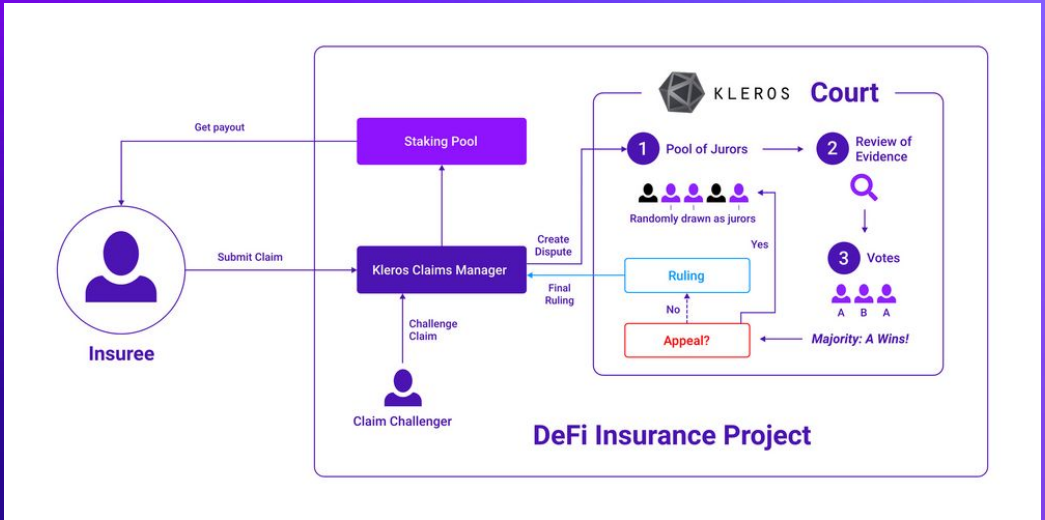
Blockchain does not have access to information about off-chain world

- ➔ **prices of assets for Defi contracts,**
- ➔ **amount of rainfall in given location for farm insurance contract, ...**


ESCROW
Home
New Invoice
New Payment
New Payment
Payment Info
Title
Eg. Marketing Services Agreement with John
Eg. Marketing Services Agreement with John
Fund Receiver
0x93ed3fbe21207ec2e8f2d3c3de6e058cb73bc04d
Enter the ETH address of the counterparty to this agreement. Make sure to use an address this party controls (Do not use an exchange address).
Amount
3
ETH
Amount that will be sent to the escrow as payment for the service. Funds will stay in the escrow until the payment is completed.
Automatic Payment (Optional)
Agreement Documents (Optional)
Extra Details | Cryptocurrency Transaction
Asset to exchange
PNK
Address to send the asset

Subjective Number Valued Questions

- partial settlements,
- insurance payouts, ...



Basic Ingredients for Oracles

- Who is participating in the oracle?
- What is the format of the votes they provide?
- How are the votes aggregated into a collective outcome?
- How are participants rewarded or penalized?

Goals for Oracles

- Speed/frequency with which returns value
- Cost
- Attack resistance - who can manipulate, by how much, and at what cost?
  - Includes goals around decentralization
- How precise (how many decimals) should information be?

Incentive Function

Is a given vote close enough to be rewarded?

If the price is ≈ 3200 USD, which of the following answers are “right”?: 3300, 3210, 3201, 3200.1, etc

Could have % accuracy requirement - e.g. within 1%

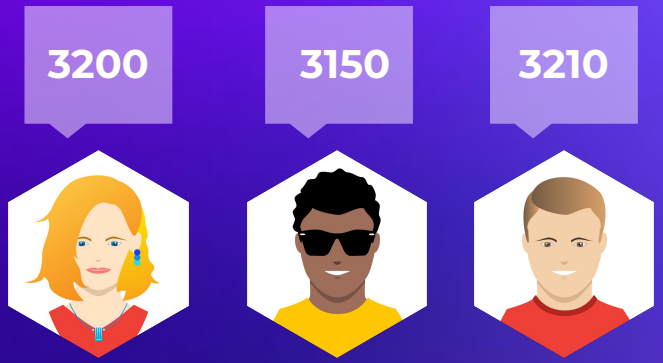
Vote Format and Aggregation

Can aggregate votes using: First past the post, ranked voting, ...

Issues with vote splitting, ...

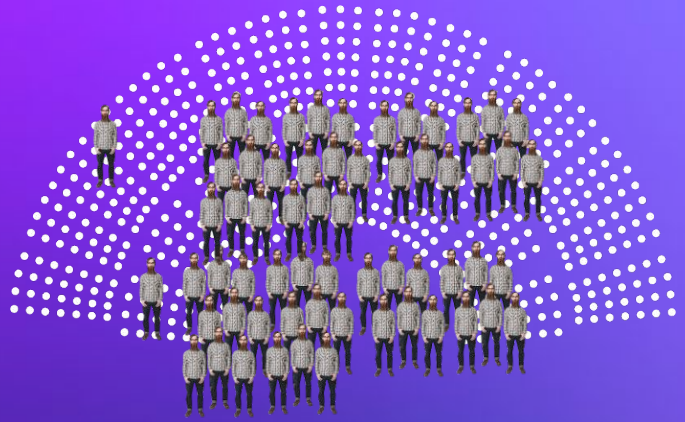
Vote Format and Aggregation

What is the price of  in USD ?

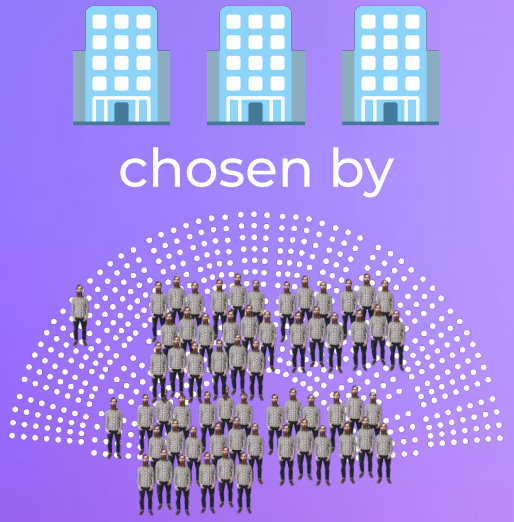


If votes are numbers, can take the median to get output value - resistant to outliers

Crowd Model vs Delegated Model



vs



Parallels to **Ethereum-style PoS** vs **Delegated Pos**

Crowd Model

- **“Crowd” of users** votes on every query
- Crowd generally needs **non-negligible amount of time** to react and vote
- Every user makes transaction → **higher gas costs**

Delegated Model

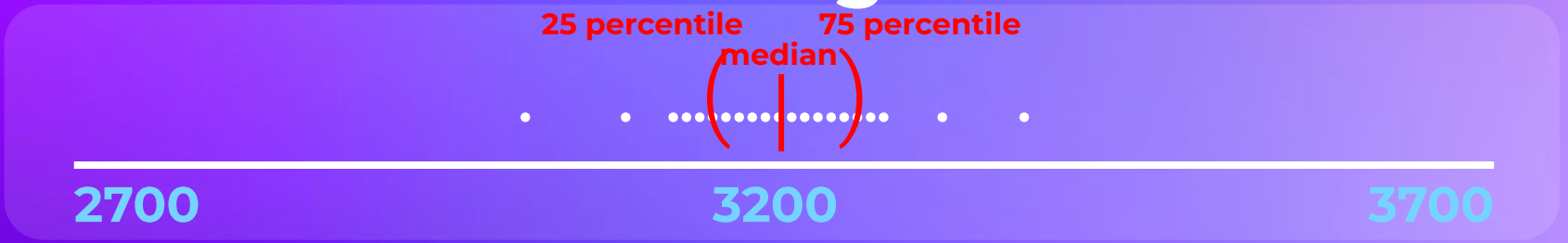
- **Designated entity** that is supposed to report
- **“Crowd” can vote out delegate** if they are doing a bad job, but do not participate in individual queries
- **Some risk of delegate abusing role** if one query more valuable than earnings can get from future participation
- Can delegate multiple entities and aggregate/take median of their answers

SchellingCoin: A Minimal-Trust Universal Data Feed

Posted by Vitalik Buterin on March 28, 2014

Research & Development

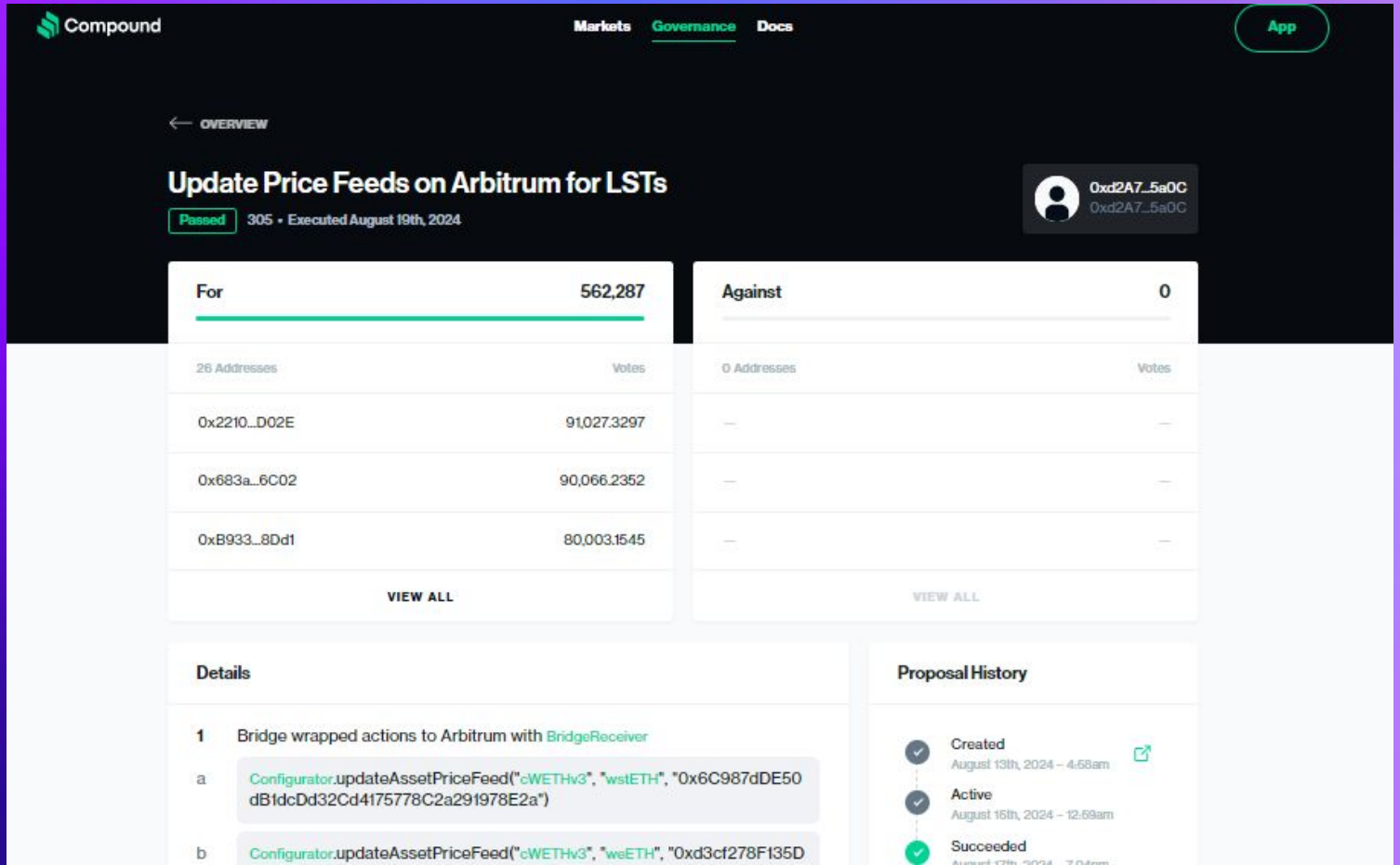
SchellingCoin



- Participants each periodically submit value
- Output the median
- Reward users in the 25-75 percentile range with N Schell tokens
- Participants not explicitly specified - supposed to be actors in PoW or PoS to achieve Sybil resistance

Chainlink

- Each node provides value - supposed to take median of a given set of reputable info sources (e.g. if price feed - Coingecko, Kaiko, etc, which themselves draw from exchange data).
- A given price feed outputs median of different node's votes
- Nodes receive payment for services.


Compound
Markets
Governance
Docs
App
OVERVIEW
Update Price Feeds on Arbitrum for LSTs
Passed
305 • Executed August 19th, 2024
0xd2A7...5a0C
0xd2A7...5a0C
For
562,287
Against
0
26 Addresses
Votes
0x2210...D02E
91,027.3297
0x683a...6C02
90,066.2352
0xB933...8Dd1
80,003.1545
VIEW ALL
0 Addresses
Votes
VIEW ALL
Details
1
Bridge wrapped actions to Arbitrum with BridgeReceiver
a
Configurator.updateAssetPriceFeed("cWETHv3", "wstETH", "0x6C987dDE50dB1dcDd32Cd4175778C2a291978E2a")
b
Configurator.updateAssetPriceFeed("cWETHv3", "weETH", "0xd3cf278F135D
Proposal History
Created
August 13th, 2024 – 4:58am
Active
August 16th, 2024 – 12:59am
Succeeded

(Partial) List of Number Oracles

Maker
(internal price oracle)

Chainlink

API3

Pyth

Kleros

UMA

Nest

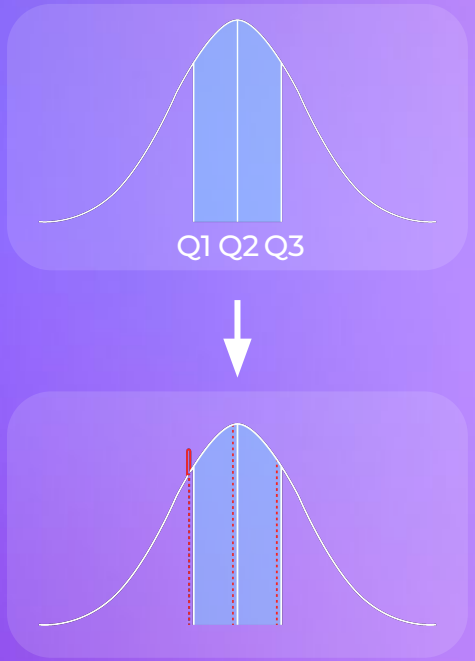
Delegated

Crowd

Micro-cheating

Another potential concern is micro-cheating. If the underlying datum is a value that frequently makes slight changes, which the price is, then if most participants in the SchellingCoin are simultaneously participants in a system that uses SchellingCoin, they may have the incentive to slightly tweak their answers in one direction, trying to keep within the 25/75 boundary but at the same time push the median up (or down) very slightly to benefit themselves. Other users will predict the presence of micro-disruption, and will thus tweak their answers in that direction themselves to try to stay within the median. Thus, if most people think that micro-cheating is possible, then mico-cheating may be possible and if they do not think so it will not be - a common result in Schelling point schemes.

Excerpt from SchellingCoin: A Minimal Trust Universal Data Feed. Vitalik Buterin. Ethereum Foundation Blog. 2014.

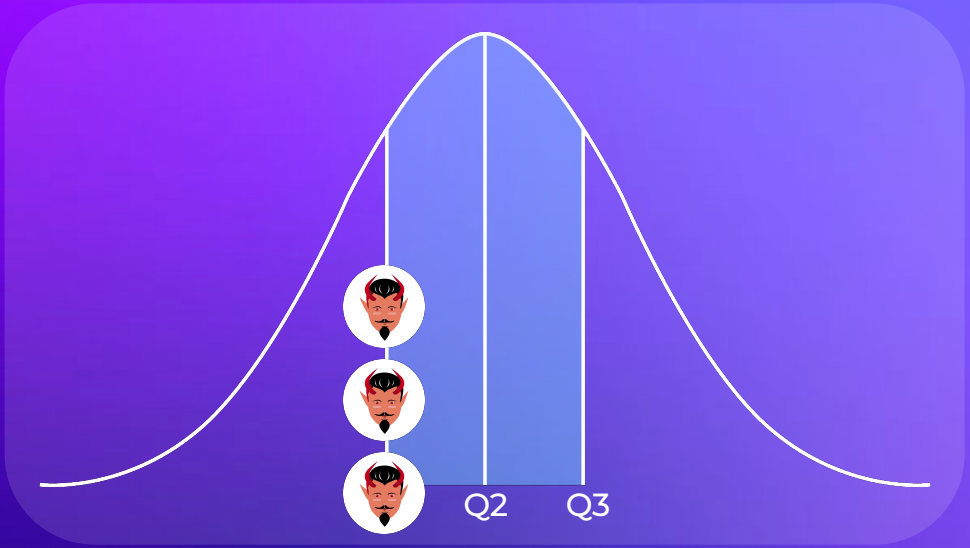


Micro-cheating

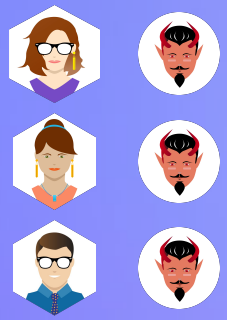
Median =
Avg of

= Spot-on

Micro-cheating



Confused OK Spot-on OK Confused



= Majority

Median = OK

Micro-cheating

Confused OK Spot-on OK Confused

= Majority

Median = Confused

Micro-cheating

- Suppose attacker has k percent of votes. How far in one direction or other can she push the median?
- Need enough votes from (confused) honest participants on edge of bell curve to get to 50%.

k+(1-k)*x=.5 -> x=(.5-k)/(1-k)

So an attacker with k percent of vote can drag the result to the

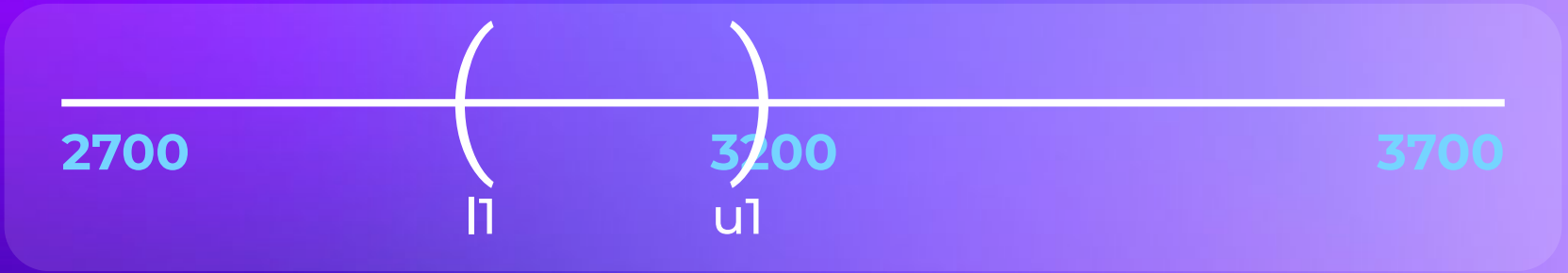
(.5-k)/(1-k)

percentile value of the distribution of honest participants.

Aggregation Rule

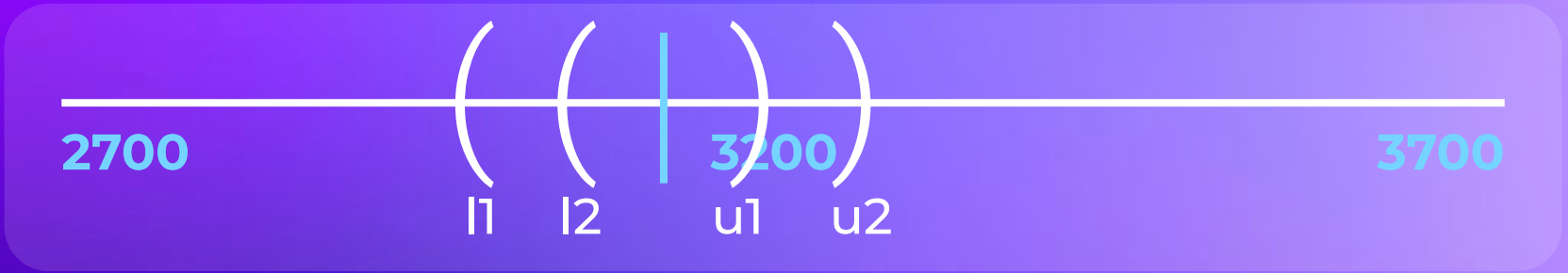
- Does it ever make sense to use anything other than taking the median?
- If voters only provide votes consisting of single numbers - probably not.

Intervals of Precision



→ Ask voters to provide lower and upper range between which confident true value lies.

Intervals of Precision



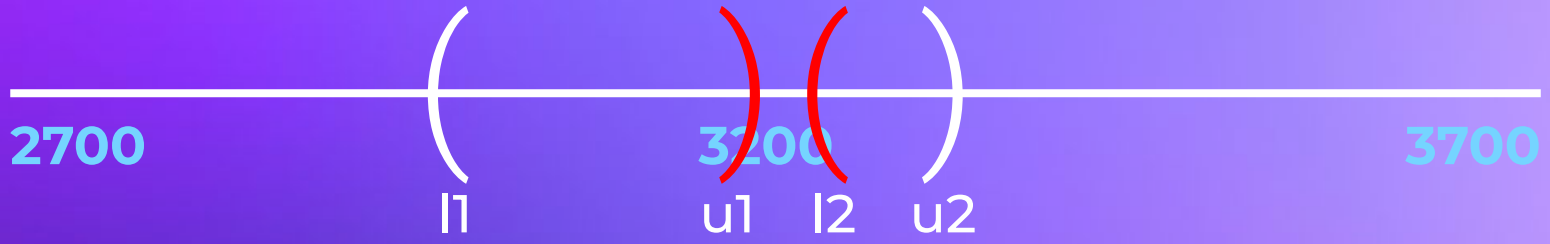
- If intervals all share a common point - compatible
  - Output central most point that is in everyone's interval

Intervals of Precision

- If intervals all share a common point - compatible
- If not - they conflict

Resolving Conflicts

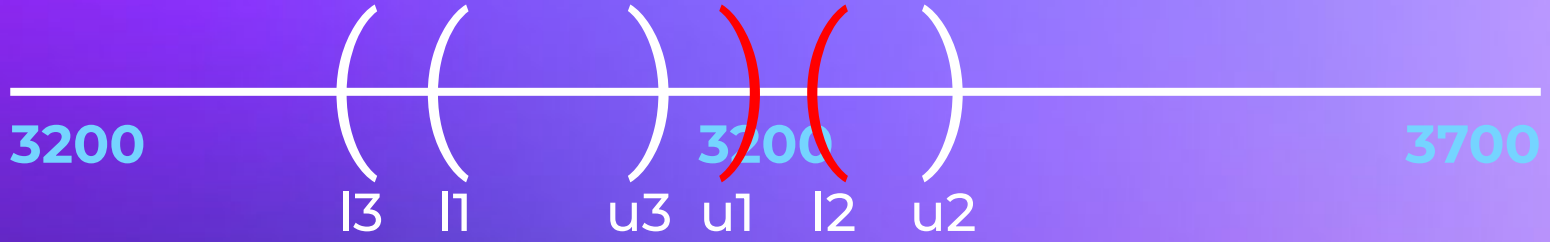
➔ At each point of conflict where have) (



Votes <	Doesn’t vote	Votes >
1		2

Resolving Conflicts

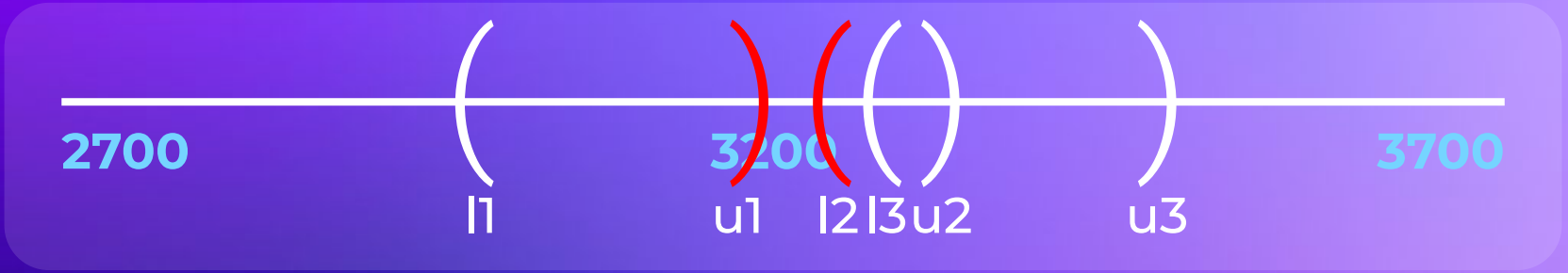
- At each point of conflict where have) (
  - If i's interval entirely to left of) (- vote <



Votes <	Doesn't vote	Votes >
1,3		2

Resolving Conflicts

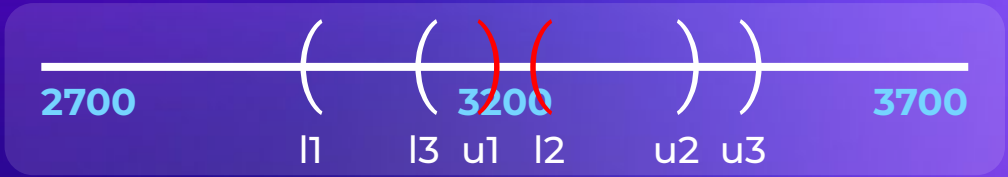
- At each point of conflict where have) (
  - If i's interval entirely to left of) (- vote <
  - If i's interval entirely to right of) (- vote >



Votes <	Doesn't vote	Votes >
1		2,3

Resolving Conflicts

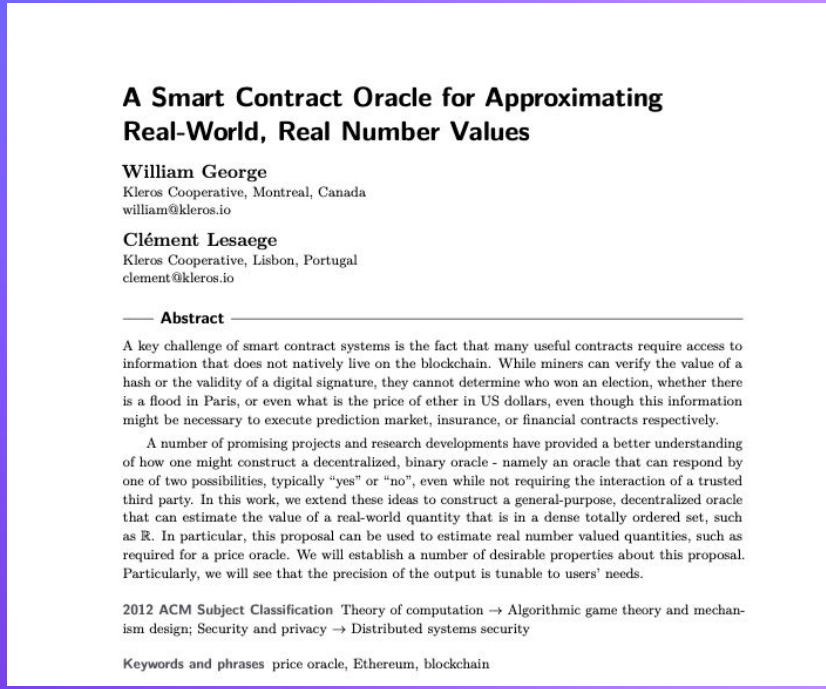
- At each point of conflict where have) (
  - If i's interval entirely to left of) (- vote <
  - If i's interval entirely to right of) (- vote >
  - If point of conflict in i's interval - don't vote
- There will be range that gets majority support from voters, output midpoint
- Can resolve conflicts in any order, will give same result



Votes <	Doesn't vote	Votes >
1	3	2

Version of this idea where ask

- ➔ separate set of actors to provide intervals,
- ➔ voters to resolve conflicts between intervals via binary search



A Smart Contract Oracle for Approximating Real-World, Real Number Values

William George
Kleros Cooperative, Montreal, Canada
william@kleros.io

Clément Lesaege
Kleros Cooperative, Lisbon, Portugal
clement@kleros.io

Abstract

A key challenge of smart contract systems is the fact that many useful contracts require access to information that does not natively live on the blockchain. While miners can verify the value of a hash or the validity of a digital signature, they cannot determine who won an election, whether there is a flood in Paris, or even what is the price of ether in US dollars, even though this information might be necessary to execute prediction market, insurance, or financial contracts respectively.

A number of promising projects and research developments have provided a better understanding of how one might construct a decentralized, binary oracle - namely an oracle that can respond by one of two possibilities, typically "yes" or "no", even while not requiring the interaction of a trusted third party. In this work, we extend these ideas to construct a general-purpose, decentralized oracle that can estimate the value of a real-world quantity that is in a dense totally ordered set, such as $\mathbb{R}$. In particular, this proposal can be used to estimate real number valued quantities, such as required for a price oracle. We will establish a number of desirable properties about this proposal. Particularly, we will see that the precision of the output is tunable to users' needs.

2012 ACM Subject Classification Theory of computation → Algorithmic game theory and mechanism design; Security and privacy → Distributed systems security

Keywords and phrases price oracle, Ethereum, blockchain

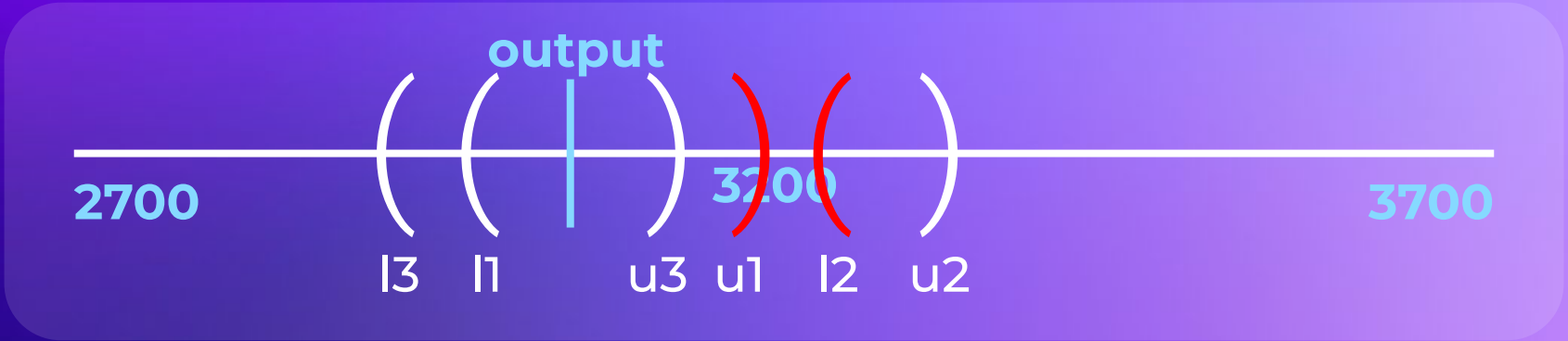
Aggregation Rule

- How does this “interval voting” compare to taking median?
  - Specifically regarding micro-cheating?

Aggregation Rule

Example vote by intervals

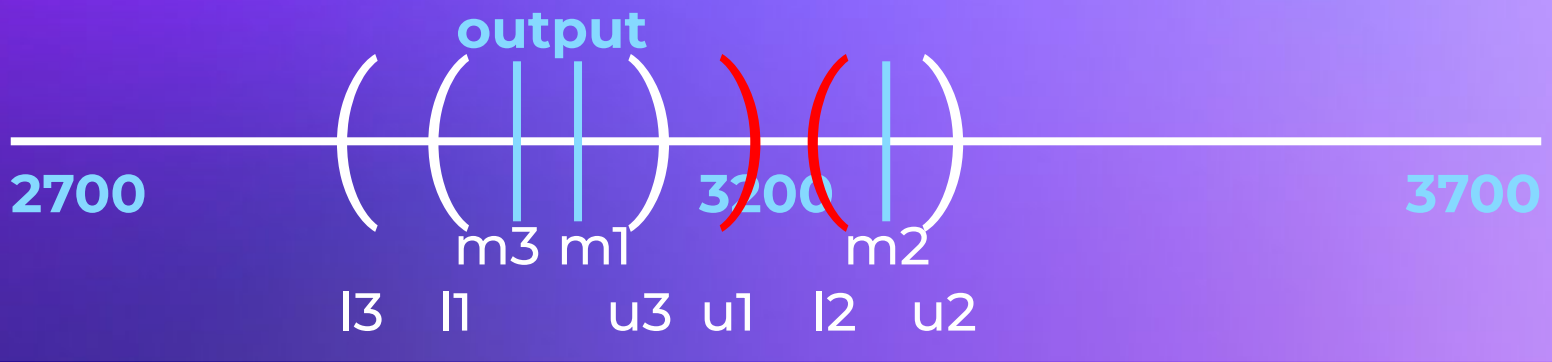
- Vote 2-1 that value less than the conflict point
- Output average of l1 and u3



Aggregation Rule

Comparison to using median

- For sake of comparison, consider each participant's vote to be midpoint of their interval
- Median of votes here is m1

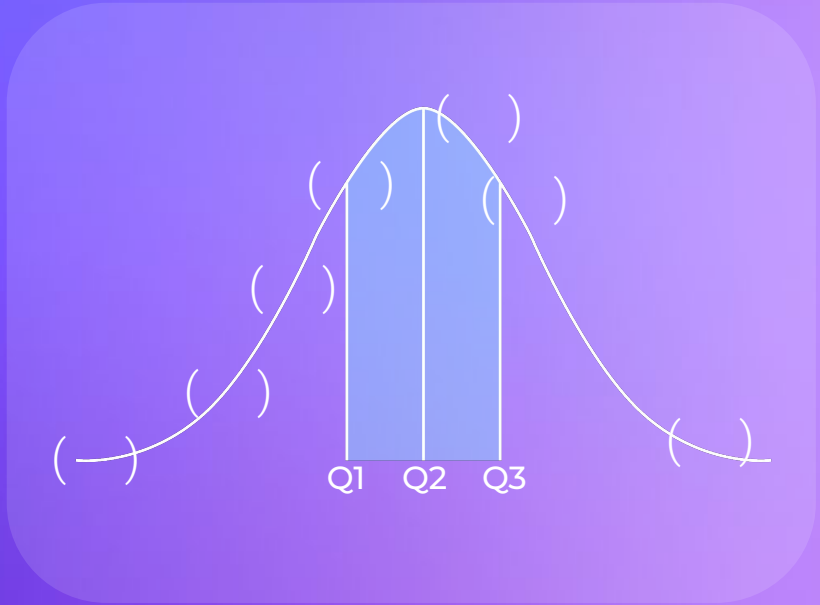


Aggregation Rule

- If participants know when they aren't experts, submit longer intervals farther from center - interval approach more robust to micro-cheating than median
- Extreme votes that would count with attacker in median approach have longer intervals, so more likely to just not count at all

Aggregation Rule

- If voters give roughly same precisions no matter how accurate they are, turns out median slightly more robust to micro-cheating than interval approach



Incentive Function

Is a given vote close enough to be rewarded?

- ➔ How much noise is there in an honest reporting of this question?
- ➔ Need a notion of distance against which to measure

Incentive Function For Interval Voting

Once v_{out} value determined by aggregation function, i's proportion of reward given by:

$$\frac{1}{2} \cdot \frac{1_{v_{out} \in (l_i, u_i)} \alpha^{-(u_i - l_i)}}{\sum_j 1_{v_{out} \in (l_j, u_j)} \alpha^{-(u_i - l_i)}} + \frac{1}{2} \cdot \frac{\beta^{\#\{c \text{ conflict point: } l_i > c, v_{out} > c \text{ or } u_i < c, v_{out} < c\}}}{\sum_j \beta^{\#\{c \text{ conflict point: } l_j > c, v_{out} > c \text{ or } u_j < c, v_{out} < c\}}}$$

- First term rewards voter for providing smaller, more precise interval (as long as it ultimately includes output value).
- Second term rewards voters for being on winning side of points of conflict.

Metrics Used in Incentives

Approach	Intuition
Reward if answer is within 1% of output	Whether close to the answer compared to the size of the answer
Reward if answer is in the 25-75 percentiles of submitted answers	Whether close to the answer compared to how other people vote
Formula from previous page for use with interval approach	Whether submit precise (short) interval compared to others + Whether vote on correct side of points of conflict

	Format of vote	Aggregation Mechanism	Metric used in Incentives
SchellingCoin	Number	Median	Closeness to the answer compared to how other people vote - notably using 25-75 rule
Chainlink	Number	Median	Subjective - matter of whether market responds by choosing a different price feed, if DAOs vote out malicious price feed
Interval approach	High and low values	Vote on conflict points by where they fall versus submitted interval	Precision of interval compared to others + Number of points of conflict where there is disagreement between voter and output

	Format of vote	Aggregation Mechanism	Metric used in Incentives
Maker **(internal price oracle)**	Number	Median	Subjective - matter of whether DAO votes out malicious data sources
API3	Number	Median	Subjective - matter of whether DAO votes out malicious data sources
Pyth	Three numbers - lower bound, estimate, upper bound	Median of lists of all three values taken together	Subjective - matter of whether DAO votes out malicious data sources

Conclusions 1

- **Basic design choice between delegated or crowd model**
  - Crowd model historically unviable for price oracles due to frequency of updates - high gas
  - With scaling solutions, maybe more viable
- **Have objective measures of micro-cheating with which to compare different models**

Conclusions 2

- **Choice of aggregation rule**
  - Median simple choice that performs well
  - Other credible choices exist
- **Choice of incentives**
  - If use delegated approach, can leave incentives to subjective decisions of how crowd chooses delegates
  - Otherwise need to build in incentives - requires notion of distance to say how far off a value is before should be penalized - lots of open research questions

Oracles for Number Values

William George
Research Lead @ Kleros

william@kleros.io
@williamhwgeorge

HOSTED BY THOMSON REUTERS INCUBATOR

European Commission | Winner of the Blockchains for Social Good Prize from the European Union's Horizon 2020 research and innovation programme

bpifrance

BLOCKCHAINGOV