Parameters specifications

We fix notation. Some of the discussed quantities are parameters that we will be able to choose. Others we can imagine as being intrinsic to a given type of dispute. The following quantities are relevant in any Kleros dispute:

- p is the probability that you correctly learn whether a submission belongs on the list or not after making an effort of cost e
- t is the larger of the percentage of entries rejected from the list and entries accepted to the list, (i.e. t is the percentage taken by the dominant outcome, $t \geq 1/2$).
- we are in an initial Kleros round with M jurors
- jurors place a deposit of d with incoherent jurors losing their deposits to coherent jurors (To be clear, d is the amount that should be lost in case of an incoherent vote, rather than the larger lockup.)
- a total of S_J is awarded to the coherent jurors in this arbitration round
- V the value at stake that could be gained by an attacker that manages to get a malicious result adopted.

Additionally, use of the appeal system requires

- y which determines the growth of appeal fees for the winner of the previous round. Namely, the winner pays $y \cdot$ (required arbitration fees) in stake, so they pay $(1+y) \cdot$ (required arbitration fees) total.
- w which determines the growth of appeal fees for the loser of the previous round. Namely, the loser pays $w \cdot$ (required arbitration fees) in stake, so they pay $(1+w) \cdot$ (required arbitration fees) total.

Whenever Kleros is used in context of a curated list, we require

- must place a deposit of deposit of $S = S_J + S_C$. Here, if the submitter loses, S_J is distributed among the coherent first-round jurors and S_C is given to the challenger
- Suppose a challenger must pay a deposit of $S' = S_J + S'_C$ and has the same possibility of determining the true result of an eventual dispute with probability p at cost e. Suppose that a proportion of u of all submissions do not belong on the list.
- K - the number of challengers we expect to participate in any given period.

1 Constraints

In this section, we list a series of relations that any choice of parameters should respect if we want Kleros to work properly. See the full fees document for further detail.

$$S_J \geq \max\left\{\frac{eM}{\left(1-(1-p)^M\right)\left(1-\frac{t}{p}\right)} - dM, \frac{(d+e)M}{1-(1-p)^M} - dM\right\}.$$

Then, the minimum amount per juror (if all M jurors vote coherently) needs to be at least

$$S_J/M \geq \max\left\{\frac{e}{\left(1-(1-p)^M\right)\left(1-\frac{t}{p}\right)} - d, \frac{(d+e)}{1-(1-p)^M} - d\right\}$$

This still depends on M, but it decreases as M grows, so it suffices to chose the payment to the juror corresponding to the M used in the first round. (Which practically should not make much of a difference, even if p is close to 1/2 and $M = 3$.)

$$\frac{e}{p} \leq S_C \leq \frac{pV - (1-p)S_J}{1-p}.$$

Note that the challengers, in fact, now pay a challenging stake and an arbitration fee stake, with the reward in case of success corresponding to the challenging stake being known in advance but the reward corresponding to the fee stake being variable. What we really need is that the expected value of the reward be greater than e/p, but for simplicity it might be a good idea here to take the constant reward for successful challenges $S_C \geq e/p$.

$$y \leq w$$

2 Trade-offs

If one chooses

$$d = e\left(\frac{1}{1-\frac{t}{p}} - 1\right),$$

that allows the lowest possible values of fees S_J such that the jurors are better incentivized to participate honestly rather than to vote randomly or to opt-out. Small decreases in d below this value require significant increases in S_J if we want the system to resist "random voting strategies/attacks", whereas increases in d above this value require only marginal increases in S_J.

Hence we probably want

$$d \geq e\left(\frac{1}{1-\frac{t}{p}} - 1\right).$$

Then, if we take d as the lower bound of this range, that choice is the most accessible and pulls in the largest possible pool of jurors. However if d is larger, that increases security against other forms of attack. Specifically, a pure $p+\epsilon$ attack in the round of M jurors, under our appeal system requires lockup costs of $(M-1)^2 d + (M-1)S_J$. So increasing d can lower the threshold at which this lockup cost becomes unviable. The lockup costs of capped/modified $p+\epsilon$ attacks and simple bribes also scale as $O(d)$. If we expect that parties will be willing to appeal to the ith round, then if we choose $d \geq \frac{V}{2^{i-1}}$, it would not be in an attacker's interest to perform a simple bribe.

Beyond the above constraints, the best choice of S_J would normally be determined by supply and demand considerations; for this, one would need some approximation of the demand and supply curves for this type of arbitration service. Conceivably we could try to compare to the fees charged by other (loser-pays) arbitration systems. One can think of $S_J \cdot (1-p)$ as an expected arbitration cost for a party that honestly believes she has a winning case.

Under the model of rational challengers and attackers, the rate of hostile submissions that go unchallenged is

$$\frac{Kr_C e + (S_J + S'_C)(1-(1-(1-p)r_C)^K)}{S_C\left(\frac{V}{S_J+S_C+V}\right) + (S_J + S'_C)(1-(1-(1-p)r_C)^K)} \cdot \frac{S_J + S_C}{S_J + S_C + V},$$

where

$$r_C = \frac{1}{p}\left(1 - \sqrt[K]{\frac{S_J + S_C}{S_J + S_C + V}}\right).$$

(In passing, r_C is the percentage of submissions that a given challenger picks to evaluate).

Here the only directly tunable variables are S_J, S_C, and S'_C (p, e, and V are thought of as being out of our control, K, and r_C are values that arise in equilibrium and respond to our choices of S_J, S_C, and S'_C). Note that

- Taking $S'_C > 0$ is counter-productive in terms of limiting this rate of errors due to lack of challenges. Even if $S'_C = 0$, challengers have to pay S_J, so the griefing factor of the time delay of a false challenge is still bounded (though if we wanted to further tune that griefing factor, increasing S'_C would certainly be a way to do so). We might want to have $S'_C > 0$ to "compensate" submitters for the annoyance of being falsely challenged, but if people want to be on the list enough, they will submit anyway. The usefulness of having a positive S'_C would seem to be mostly psychological.
- This relationship is complex, as if one takes S_C to be large, you would think that that would encourage a high level of challenging, but then the

rate of (rational) attackers would be expected to drop, so there may be so few attacks that it is not worth the time of challengers to search for them. However, for realistic choices, increasing S_C causes this error rate to go down, so the question is finding an acceptable balance between the deposit required to submit to the list and the acceptable error rate. Specifically the upper bound above on S_C given in terms of p, V, and S_J should be considered a hard limit as beyond that point the expected value of submitting to the list is negative.

In choosing y and w, the tradeoffs to consider are essentially that if y and w are large, insurers will be willing to evaluate cases and participate after a relatively small number of appeals. If y and w are small, insurers will take more rounds to participate, but the required fees will be smaller and more accessible. Particularly, if

- If w is large but not y, then insurers will have an incentive to evaluate the case of previous round winners once the previous round losers fees are paid; this increases resistance to bank attacks but not other kinds of attacks.
- If $y \leq w$ are both large, insurers will have an incentive to evaluate cases at the beginning of the appeal period, increasing resistance to all kinds of attacks.

As a first, rather rough approximation of how y and w determine willingness to appeal, suppose Alice won the previous round and Bob lost the previous round. Then, the insurer will be willing to insure Alice if he thinks that Alice will win the overwhelming majority of any future appeal rounds *and* eventually win the dispute with probability of winning p_A (expressed as a fraction between 0 and 1) such that

$$p_A \geq \frac{1}{\frac{w}{y+w}+1}.$$

On the other hand, in any case, the insurer is willing to fund either party if the insurer believe that that party will win with probability at least

$$\frac{1}{\frac{y}{y+w}+1}.$$

More detailed, less approximative bounds (that consider various scenarios for the results of intermediate appeal rounds) can be found in the full, technical fees document. Then the tradeoff to take here is setting y and w such that these cutoffs are small enough that we think the appeal system is likely to attract insurers, while y and w are still small enough so that appeal costs are nevertheless seen as reasonable.

Finally, if

$$w \cdot 2^i \geq \frac{e}{p}$$

in the ith appeal round, insurers will have an incentive to evaluate (a randomly chosen percentage of) the appealed cases, at least on behalf of the previous round loser after the previous round loser has paid their fees. The percentage of honest cases that fail to get evaluated can be analyzed with similar reasoning to the situation for challengers. Note that if i is large enough, this inequality is satisfied, but w influences at what round insurers begin to consider cases and at what rate.