$p + \epsilon$ and counter-coordination comments

August 26, 2018

1 Simple counter-coordination (CC)

The counter coordination described here is compatible with public, non-commit and reveal vote model/absence of anti-pre-relevation mechanisms.

Parameters:

- Deposit D (which will be larger than the deposit d for the main game),
- Margin of error E,
- Target margin of victory F (will require $F \leq E$),
- L which will is the maximum amount the coherence payment can take in the primary game while still having all the participants in the counter coordination guaranteed to receive the same total payoffs between the main game and the counter-coordination,
- contract address and
- disputeID for dispute on which to counter-coordinate,
- outcome that attacker is trying to obtain (again denote by Y, denote the other outcome by X),
- ϵ.

Remark 1. *In possibly more sophisticated versions, don't assume that the contract knows epsilon or which option that the attacker is trying to obtain. Perform straw vote to determine which side to coordinate on. Ideally would create a de-facto vote where jurors have an incentive (due to the funds from the attacker) to vote in a straw vote that determines with great likelihood the result of the actual vote and has incentives in itself to vote coherently. Then conceivably an attacker might launch a $p + \epsilon$ attack against this straw vote at the expense of greater risk if the attack fails and larger capital lock-up.*

1. Pull period lengths from relevant dispute

   During the voting phase of the relevant dispute.

2. Jurors submit deposit of D to cc contract.

3. Let M be the total number of jurors ruling on the case. Let S be the number of jurors participating in the counter-coordination. If $S \geq M/2 + E$ (namely if S is large enough to force a particular ruling on the case even if E jurors vote in the other direction) continue, if not abort and release deposits to participants.

4. Use an (upredictable) RNG to choose $\lceil M/2 \rceil + F$ from among the S participants to vote X in the dispute. The remaining $S - \lceil M/2 \rceil - F$ participants are instructed to vote Y.

   Notation: Denote the number of counter-coordinators who are ordered to vote X and do so by: x. Similarly denote the number of counter-coordinators who are ordered to vote Y and do so by: y. Denote the total number of jurors who signed up for the counter-coordination but defected and did not follow their required vote by: z. (Hence $x + y + z = S$.)

   During the execution period, once the final results are known:

5. In the extremely unlikely case that $S = z$, i.e. everyone defected which in particular implies that Y won the case, there should be some error thrown, for example that burns the deposits.

6. If X wins the dispute (which will be the case as long as no more than F of the CC participants that are ordered to vote X defect and vote Y), then including the bribes paid to the Y voters, the only difference between the payoffs to the CC participants is the ϵ. Any CC participants who deviate from ordered vote lose the deposit D. Pay to each juror who followed their instruction to vote X:

   $$D + D \cdot \frac{z}{S - z} + \epsilon \cdot \frac{y}{S - z}.$$

   Pay to each Y voter

   $$D + D \cdot \frac{z}{S - z} + \epsilon \cdot \left(\frac{y}{S - z} - 1 \right).$$

7. If Y wins dispute, the contract attempts to even out the payments to the participants and distribute lost deposits from defectors to cover/minimize the losses of the non-defecting participants. Calculate the sum of

   $$B = \min \{\text{total payouts paid to Y CC voters}, L \cdot y\} - d \cdot x + D \cdot z.$$

   (Note that B is defined even if the no one in this round of the dispute voted for Y and the coherence payout is undefined, the lost deposits going to the governor, as in this case their are no Y CC voters so the total paid to Y CC voters is 0. Additionally, if there are any non-defecting Y voters then in particular the coherence payout will be defined.)

Pay to each non-defecting X voter

$$D + d + \frac{B}{S - z}.$$

Pay to each non-defecting Y voter

$$D - \min\{\text{coherence payment}, L\} + \frac{B}{S - z}.$$

Remark 2. *If there was only a single round that determined the final result, F could be set so that CC participants are guaranteed to at least break even; namely that B is positive. However, it is possible that a counter-coordination against a given round has no defectors, yet that a later round overrule them, producing an opposite final result.*

Notable limitation: it is in the best interest of an individual juror that the other jurors counter-coordinate and that X wins the dispute while the juror does not participate in the counter-coordination and votes Y, obtaining the full return of the $p + \epsilon$ bribe. A perfectly rational juror might only participate in the CC if she thought her participation would make a difference in whether it activates or not, which as the total deposits the CC contract has received is rather visible. That said, the mere threat that a CC might be viable may discourage potential attackers from launching $p + \epsilon$ attacks.

Proposition 1. *The contract has enough in deposits to make the required payouts in any eventuality if:*

$$D \geq \max\{\epsilon, L + d\}.$$

Proof. Note that we cannot make any assumptions about the vote counts in the round of the counter-coordination that depend on a count of how many counter-coordinators defect because the result of the dispute could be overruled in a later appeal round.

For the payments in step 6, a total of

$$(S - z)\left(D + D\frac{z}{S - z} + \epsilon\frac{y}{S - z}\right) - \epsilon \cdot y$$

must be paid out. However, this simplifies to SD, which is the amount input in deposits.

The payment in step 6 to jurors who follow their instructions to vote X is clearly positive. For the payment for voters who follow instructions to vote Y,

$$D + D \cdot \frac{z}{S - z} + \epsilon \cdot \left(\frac{y}{S - z} - 1\right),$$

the middle term is positive regardless of the number of defectors assuming that $S \neq z$. Also

$$\left(\frac{y}{S-z} - 1\right) \geq -1.$$

Then,

$$D + \epsilon \cdot \left(\frac{y}{S-z} - 1\right) \geq D - \epsilon \geq 0,$$

as we have assumed $D \geq \epsilon$.

Hence, for the step 6 payments, we are making a set of positive payments that sum up to what has been deposited; hence a sufficient amount has been deposited.

For the step 7 payments, the sum of payments is

$$(S-z)\left(D + \frac{B}{S-z}\right) + x \cdot d - y \cdot \min\{\text{coherence payment}, L\}$$

$$= (S-z) \cdot D + B + x \cdot d - y \cdot \min\{\text{coherence payment}, L\}$$
$$= SD.$$

Suppose that there is no non-defecting Y voter. Then the amount in paid to each non-defecting X voter is clearly the amount in the contract divided by the number of non-defecting X jurors.

Hence, we can suppose that there is at least one non-defecting Y voter. Then the coherence payout is defined, and as

$$D + d + \frac{B}{S-z} \geq D - \min\{\text{coherence payment}, L\} + \frac{B}{S-z},$$

it suffices to show that

$$D - \min\{\text{coherence payment}, L\} + \frac{B}{S-z} \geq 0.$$

Then $B \geq -dx$ and $S - z = x + y \geq x$. So

$$\frac{B}{S-z} \geq -d.$$

Thus as $D \geq L + d$, this condition holds.

□

We briefly examine under what conditions a counter-coordinator would have an incentive to defect from being instructed to vote from X to instead voting Y. There are three possible situations:

1. The defection does not change the outcome and the outcome is X wins.
2. The defection does not change the outcome and the outcome is Y wins.

3. The defection changes the result from X winning to Y winning.

If

$$D \geq \max\{2\epsilon, L + d\},$$

then the calculations in Proposition 1 show that the payouts to the non-defecting counter-coordinators are not only all non-negative, but they are at least ϵ when the counter-coordination succeeds. In particular, (as a $p + \epsilon$ bribe pays what one *would have* gotten had they coherently voted X plus an additional ϵ), if the counter-coordination succeeds and X ultimately wins, a counter-coordinator receives no less from the counter-coordination contract for voting as instructed than she stands to gain by taking the bribe and defecting to Y. So in case 1 it is no worse to not defect.

Note that if L is the same as the maximum coherence payout, then the amounts that counter-coordinators receive including both the counter-coordination contract and the main dispute would be the same regardless of whether they are ordered to vote X or Y. However, the maximum coherence payout is $(M-1)d$ which is paid when a single juror rules Y in a given round but then Y ultimately wins in appeal. This growth of the potential coherence payouts makes $p + \epsilon$ attacks more expensive and risky for attackers, but using $L = (M-1)d$ would require $D \geq Md$ which is unviable for large M.

Namely, if $L \geq (M-1)d$,

$$\text{Global payoff}_{\mathcal{USR}\text{ does not defect}}(\text{selected to vote } X, \text{ votes } X)$$

$$= \text{Global payoff}_{\mathcal{USR}\text{does not defect}}(\text{selected to vote } Y, \text{ votes } Y).$$

As $\mathcal{USR}$ defecting increases the number of Y votes and decreases the number of X votes compared to not defecting, the coherence payout paid to the Y voters decreases if $\mathcal{USR}$ defects. Hence

$$\text{Global payoff}_{\mathcal{USR}\text{ defects}}(\text{selected to vote } Y, \text{ votes } Y)$$

$$\leq \text{Global payoff}_{\mathcal{USR}\text{ does not defect}}(\text{selected to vote } Y, \text{ votes } Y).$$

As the payouts from the counter-coordination contract are all non-negative, counter-coordinators who are instructed to vote Y and then do so are at least as well off as counter-coordinators who are selected to X and defect.

$$\text{Global payoff}_{\mathcal{USR}\text{ defects}}(\text{selected to vote } X, \text{ votes } Y)$$

$$\leq \text{Global payoff}_{\mathcal{USR}\text{ defects}}(\text{selected to vote } Y, \text{ votes } Y).$$

So, putting this together

$$\text{Global payoff}_{\mathcal{USR}\text{ defects}}(\text{selected to vote } X, \text{ votes } Y)$$

$$\leq \text{Global payoff}_{\mathcal{USR}\text{ does not defect}}(\text{selected to vote } X, \text{ votes } X).$$

Hence it is not in the juror's interest to defect in case 2.

The global payoff for an X voter if X wins is

$$\text{coherence payment} + D + D \cdot \frac{z}{S-z} + \epsilon \cdot \frac{y}{S-z},$$

whereas the payout for an X if Y wins with the same vote counts (for example, if the decision is reversed in appeal) is

$$D + d + \frac{B}{S-z} - d = D + \frac{B}{S-z}.$$

However,

$$\text{coherence payment} + D \cdot \frac{z}{S-z} \geq \frac{B}{S-z}$$

$$\Leftrightarrow (S-z) \cdot \text{coherence payment} + Dz \geq y \cdot \text{coherence payment} - dx + Dz$$

$$\Leftrightarrow x \cdot \text{coherence payment} + dx \geq 0.$$

So the payoff for a user for not defecting and voting X in case 3 is even higher than that in case 2, while the payoffs for defecting and voting Y are the same in these two cases.

Thus, a user has no internal incentive to defect from the counter-coordination and vote Y as long as $D \geq \max\{2\epsilon, L+d\}$ and $L \geq (M-1)d$. External incentives (such as additional bribes from the attacker) may of course still cause defections.

To come/to be filled in (not necessarily relevant for the Doge pilot), argument that considers mixed strategies where counter-coordinators selected to vote X defect with some probability to Y (they want to be one of few Y voters in a round but for Y to win in appeal to get the largest possible coherence payout), bound the probability that more than F defections happen (with rational participants playing such a mixed strategy) in terms of L.

For the Doge pilot, the M's are sufficiently small that paying a deposit of $(M-1)d$ is still workable, and in any event the margins E and F can only take a limited number of integer values. So for the first ruling we pretty much must take $E = 1$, $F = 0$, $L = 2d$. Namely, the counter-coordination contract requires all three jurors to commit in order to activate, one of whom is told to vote Y and the other two X.

For the first appeal, the two reasonable sets of parameters seem to be

- $E = 2$, $F = 1$ - namely that at least six of the seven jurors must commit and five are instructed to vote X, with other(s) voting Y with a margin of error for a single defector or
- $E = 1$, $F = 0$ - at least five of the seven jurors must commit and four are instructed to vote X, the other(s) Y with no margin of error for defection.

Here, $L = 6 \cdot 200 = 1200$ PNK is probably not prohibitive. Then we take $D \geq \max\{2\epsilon, L+d\}$.