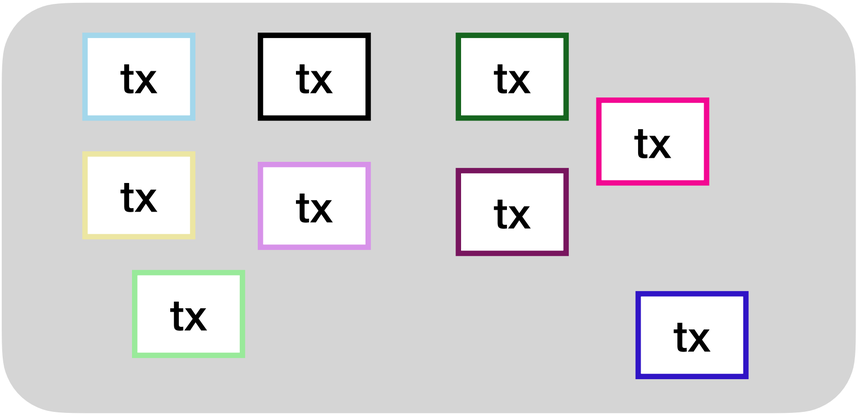
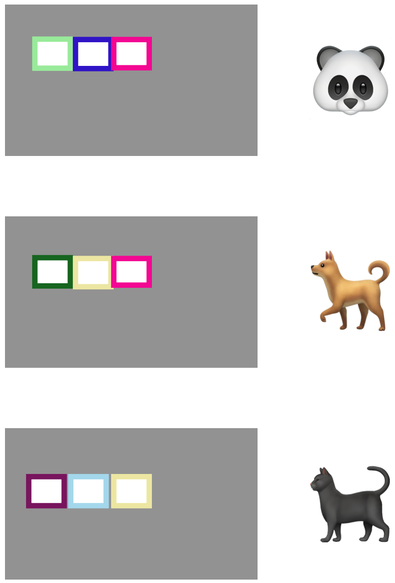
Rational Censorship Attack: Breaking Blockchain with a Blackboard

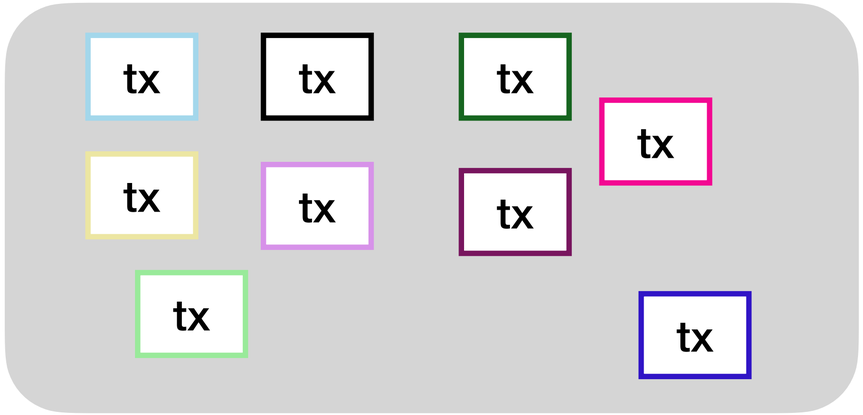
Michelle Yeo, Haoqian Zhang

Blockchain

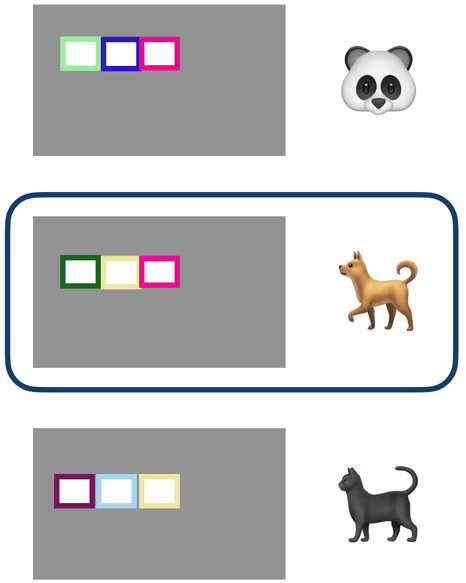




Blockchain



Consensus



Censorship resilience

- Fundamental blockchain property!
- Two levels:
  - Transaction level censorship
  - Consensus level censorship

Censorship resilience

- Fundamental blockchain property!
- Two levels:
  - Transaction level censorship
  - Consensus level censorship

Consensus level censorship is costly…

- Need sufficient resources
  - 51% mining power in PoW
  - 34% stake in PoS
- Costly to coordinate
- Irrational in the wider ecosystem
  - Short-term benefit of attack outweighed by long-term cost of drop in currency

Consensus level censorship is costly…

- Need sufficient resources
  - 51% mining power in PoW
  - 34% stake in PoS
- Costly to coordinate
- Irrational in the wider ecosystem
  - Short-term benefit of attack outweighed by long-term cost of drop in currency

… or is it?

Previous work

Rational attacks on blockchains

[FB19], [ZBMEF24]

Consensus-level censorship

[Miller13], [Bonneau16], [MHM18], [AKLM24]

Previous work

Rational attacks on blockchains

[FB19], [ZBMEF24]

Consensus-level censorship

[Miller13], [Bonneau16], [MHM18], [AKLM24]

This work: first consensus-level censorship attack

- No coordination needed
- Rational for nodes to participate
- Dominant strategy incentive compatible (truthfulness is incentive-compatible)

Blockchain model

- $n :=$ total users
- $v_i :=$ voting power (i.e., proportion of resource) of ith user
  - $\sum_{i=1}^{n} v_i = 1$
- Block reward normalised to 1
  - Expected reward for each user for each unfinalised block is v_i

Cost model

- Main cost: downward movement of cryptocurrency price if attack is detected
- Detection threshold η represents some threshold of users excluded such that attack becomes detectable if more than η users excluded

$$f(\sigma) = \begin{cases} 0, & \text{if } k < \eta \\ \alpha & \text{if } k \geq \eta \end{cases}$$

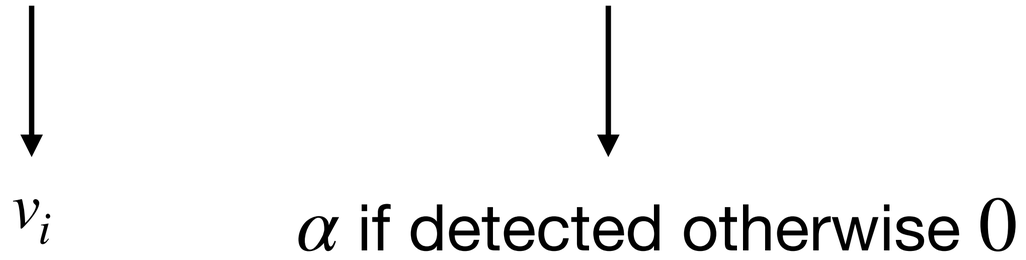
undetectable setting

detectable setting

Cost of strategy that excludes k users

Adversarial model

- All participants rational!
- Expected payoff: average reward less incurred costs



Rational censorship attack assumptions

1. (Honest threshold.) Let $\mathcal{N}_h$ represent honest nodes and if $sum_{i \in \mathcal{N}_h} v_i > t$ for some threshold $t \geq \frac{1}{2}$ then system functions perfectly

2. (Random and unknown response order.) The order in which nodes respond to attack is unknown.

3. (No big player.) $\nexists i \,|\, v_i > (1 - t)$

4. (Known detectability threshold.) η is known to all participants. Also detectability cost α is larger than any profits gained from attack

Setup phase

- Post “call to attack” (CTA) smart contract on blockchain
- Wait for other nodes to respond sequentially with their declared powers
- Setup phase concludes with success if sufficiently many nodes with cumulative power $> t$
- Attack coalition: first $n' < n$ nodes that responded s.t. their cumulative power $> t$

Algorithm 1: CTA smart contract

input: node powers v_i, node ids x_i

$v :=$ declared power of user who posted the smart contract
$V := v$
$\mathcal{N}_a := \emptyset$
ActiveAttack $:= 0$
while *new pair* $\{x_i, v_i\}$ *and* $V < t$ *and* $n - |\mathcal{N}_a| < \eta$ *and clock* $< T$ **do**
  $V := V + v_i$
  $\mathcal{N}_a := \mathcal{N}_A \cup x_i$
end
if $V < t$ *and clock* $\geq T$ **then**
  abort
end
return ActiveAttack $= 1$

Attack phase

- Nodes in the attacking coalition censor messages from nodes that are outside the coalition
- $V :=$ sum of powers of nodes in attack coalition
- Effect of successful attack: user i in the attack coalition has expected reward of $\frac{v_i}{V} \geq v_i$ as $V \leq 1$

Algorithm 2: Rational Censorship Attack

Send *Upon receiving a send request m to i:*
  if *attack is not active or $i \in \mathcal{N}_a$* **then**
    Send m to i
  end

Deliver *Upon receiving a deliver request m from i:*
  if *attack is not active or $i \in \mathcal{N}_a$* **then**
    Deliver m from i
  end

Analysis

Theorem (informal)

Upon seeing the CTA smart contract, it is rational for the remaining users to respond to agree to be part of the attack coalition and declare their powers truthfully.

Analysis

- Main idea of proof:

1. Joining the attack only improves expected utility!
   - Detectability threshold is known
   - Cost of “failed” attack is just the same utility as if attack did not happen
2. Random and unknown ordering of players enforces truthful declaration of powers
   - Declaring too much: attack might not succeed
   - Declaring too little: attack coalition might be larger so less payoff

Potential countermeasures

- Modify block reward such that it scales with the number of participants
  - Censorship becomes less profitable
- Launch multiple attack smart contracts simultaneously
  - Coordination (to join a specific coalition) is costly

Note:

The countermeasures make attack less profitable/more costly but do not fully prevent it!

Why are such attacks unobserved?

- (Undetectability.) Attacks only censor a few nodes, thus undetectable
- (Irrationality.) Nodes are not entirely rational
- (Undeniable evidence.) Public nature of smart contracts makes attack undeniable

Conclusion

- Rational censorship attack: low cost attack on blockchain systems
- Incentive compatible to join attack and declare powers truthfully
- Future work:
  - Stronger countermeasures (using cryptography?)
  - Comparison (social welfare) of our attack strategy to other equilibrium strategies

michellexyeo@gmail.com