HOW TO DESTROY BITCOIN: GAME THEORY AND ATTACKS

Nadir Akhtar
Aparna Krishnan

BLOCKCHAIN
AT BERKELEY

LECTURE OVERVIEW

1. POOL STRATEGIES
2. FORKING & DOUBLE SPENDING
3. CENSORSHIP
4. SELFISH MINING
5. DEFENSES

1 POOL STRATEGIES

POOL REWARDS

BASIC POOL REWARD SCHEMES

Pay-per-share

Pool pays out **at every share submitted.** By default will be proportional to work done by individuals

1. More beneficial for **miners**
2. Individual miners have no risk from reward variance
   a. Pool takes on the risk completely
3. Problem: No incentive for individuals to actually submit valid blocks
   a. Individuals are paid regardless

Proportional

Pool pays out **when blocks are found,** proportional to the work individuals have submitted for this block

1. More beneficial for the **pool**
2. Individual miners still bear some risk in variance proportional to size of the pool
   a. Not a problem if pool is sufficiently large
3. Lower risk for pool operators - only pay out when reward is found
   a. Individuals thus incentivized to submit valid blocks

POOL REWARDS

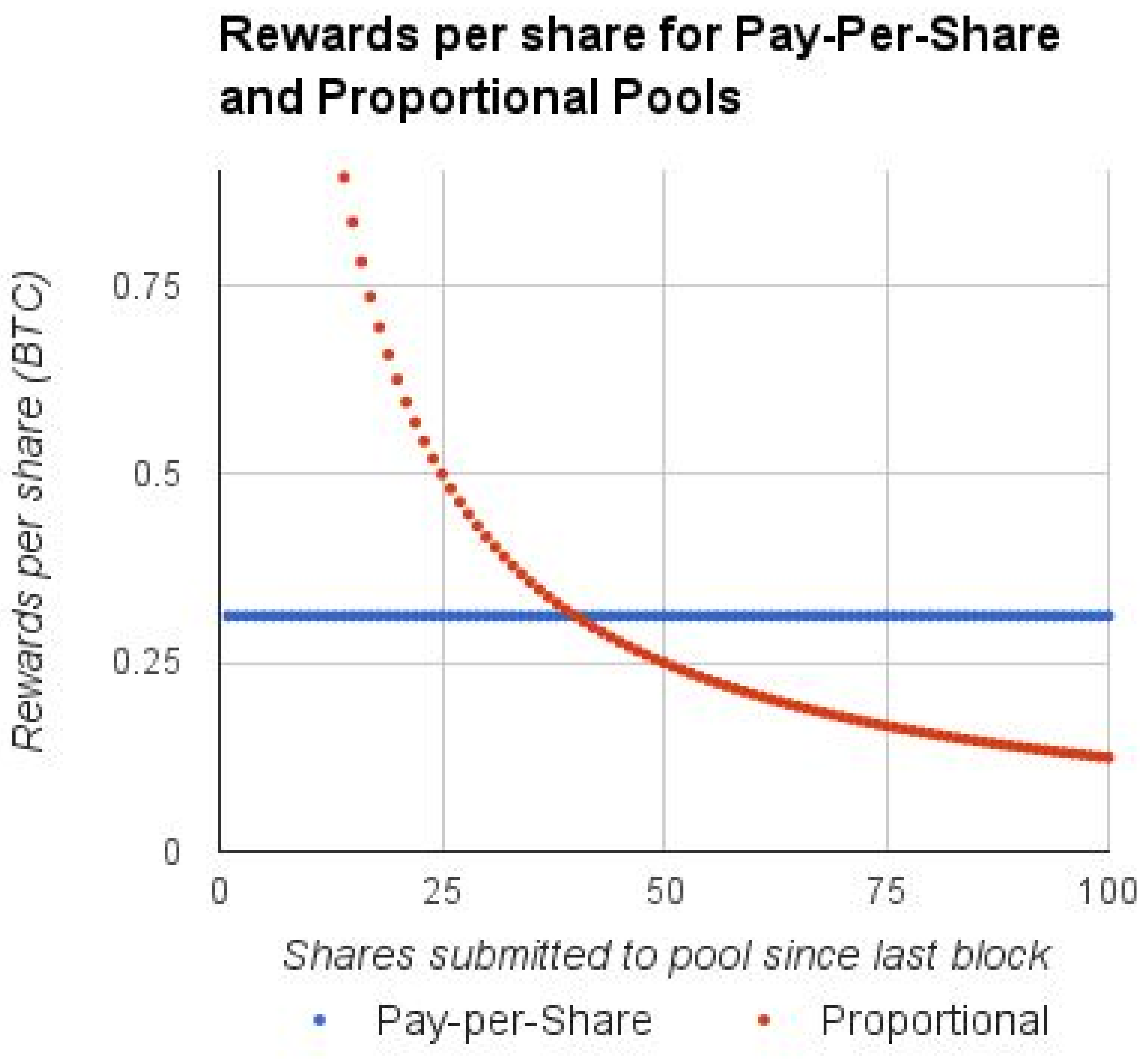
POOL HOPPING

Pool hopping: switching between pools to increase total rewards

- Proportional pool pays larger amount per share if a block is found quickly

Example clever strategy:

- Mine at **proportional** pool shortly after a block was found (while rewards are high)
- Switch to **pay-per-share** pool when once proportional pool is less profitable



Parameters:

- Pool has 10% of network hashrate
- 4 shares expected per valid block

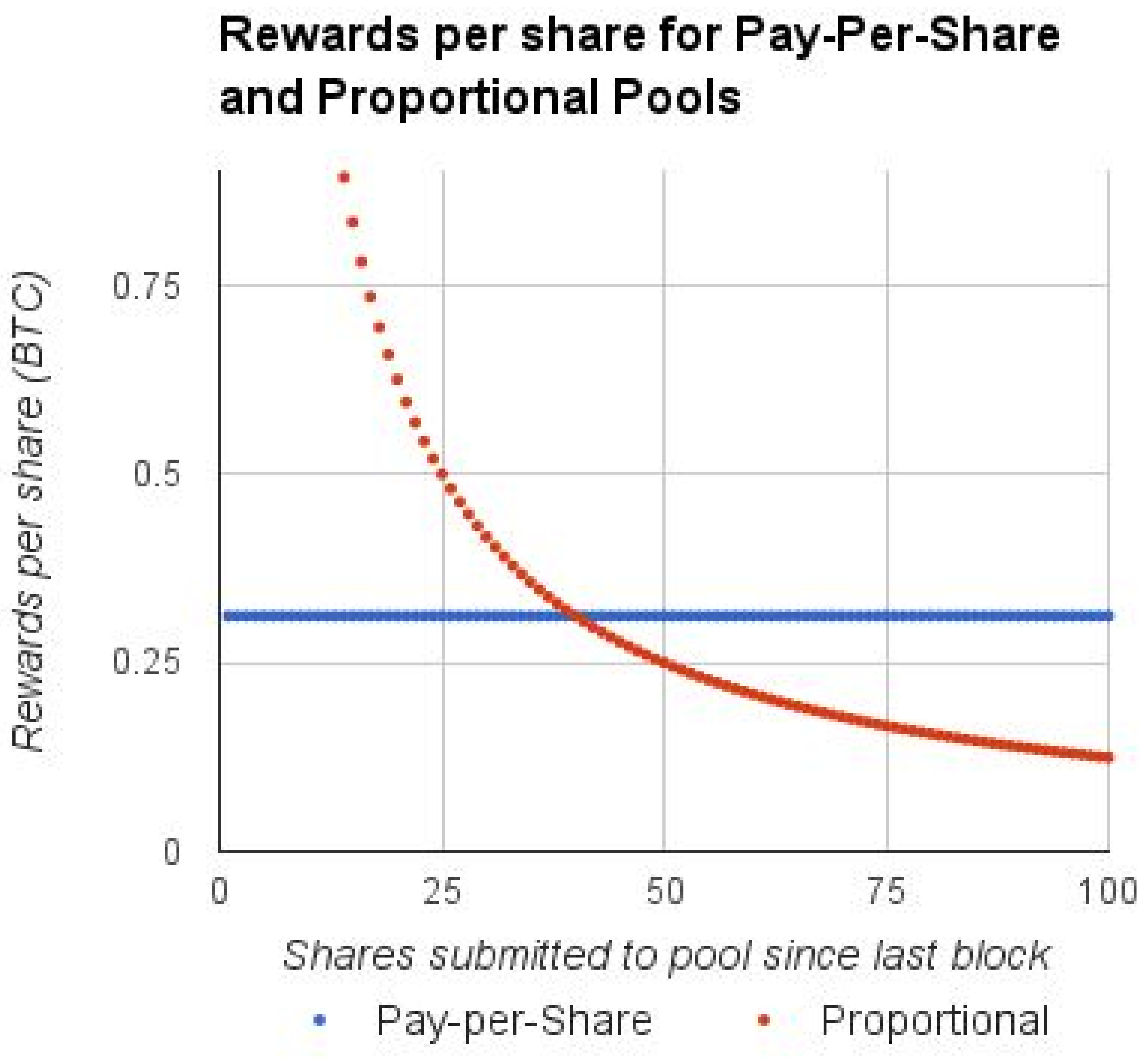
BLOCKCHAIN AT BERKELEY

POOL REWARDS
POOL HOPPING

Therefore, proportional pools are **not feasible in practice**

- Honest miners who stay loyal to one pool are cheated out of their money

Designing a mining pool reward scheme with aligned incentives that is not vulnerable to pool hopping remains an **open problem**



Parameters:

- Pool has 10% of network hashrate
- 4 shares expected per valid block

POOL CANNIBALIZATION

NOM NOM NOM NOM

Cannibalizing Pools - Distribute some small % of mining power equally among all other pools, withhold valid blocks.

- Rewards will still be received
- Undetectable unless statistically significant

Image source:
http://cdn3-www.craveonline.com/assets/uploads/2014/08/Hannibal-Cookbook.jpg

POOL CANNIBALIZATION

EXAMPLE

Givens:

- You have 30% of the hashrate. Assume 1 BTC block reward. All of the following numbers are expected value.
- 30% HR (hashrate)
  - = 30% MR (Mining Reward)
  - = 0.3 BTC

You buy more mining equipment, worth 1% of current network hashrate

Standard mining strategy:

- Add 1% HR => 31/101 = 30.69% HR = .3069 BTC
  - Revenue gain = **0.0069 BTC for 1% hashrate added**

Pool cannibalizing strategy:

Other pool hashrate breakdown:

- (70/71 honest, 1/71 dishonest) = 70% honest **effective** hashrate = .7 BTC
- You own (1/71) of other pools, so expected value of mining there is (1/71) * .7 = 0.0098 BTC
- **0.0098 (cheat) > 0.0069 (honest)**

More profitable to cannibalize pools than mine honestly

POOL WARS

THE GAME OF LIFE

- Attack decisions resemble an iterative game
    - Two players: pool A and pool B
- Each iteration of the game is a case of the Prisoner's Dilemma
    - Choose between attacking or not attacking

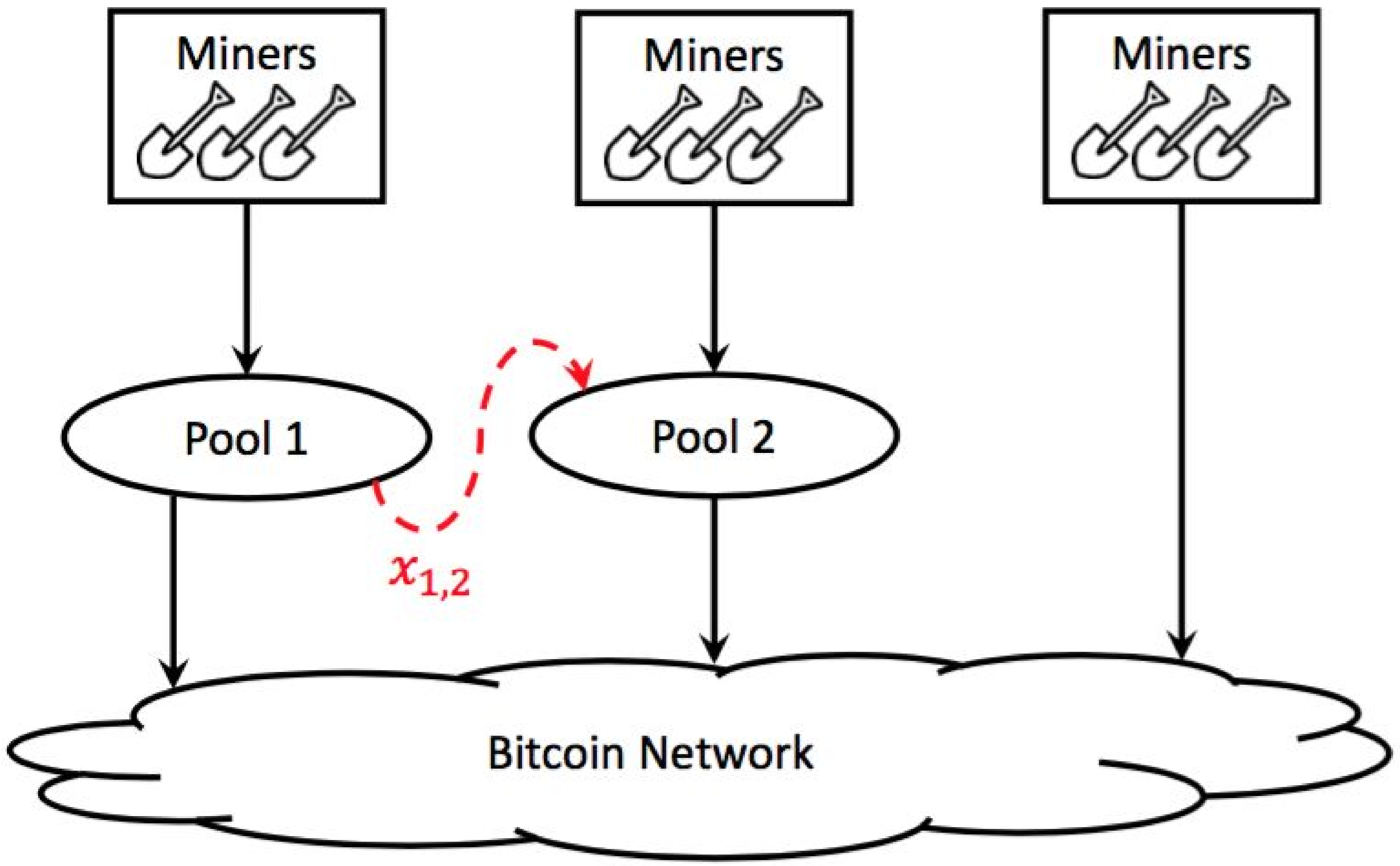


Fig. 3. The one-attacker scenario. Pool 1 attacks pool 2.

POOL WARS

NASH EQUILIBRIUM

- If pool A chooses to attack pool B, pool A gains revenue, pool B loses revenue
  - Pool B can retaliate by attacking pool A and gaining more revenue
- Thus, attacking is the dominant strategy in each iteration
  - Therefore if both pool A and pool B attack each other, they will be at a Nash Equilibrium
  - **Both will earn less than they would have if neither of them attacked.**

Pool 2 \ Pool 1	no attack	attack
no attack	$(r_1 = 1, r_2 = 1)$	$(r_1 > 1, r_2 = \tilde{r}_2 < 1)$
attack	$(r_1 = \tilde{r}_1 < 1, r_2 > 1)$	$(\tilde{r}_1 < r_1 < 1, \tilde{r}_2 < r_2 < 1)$

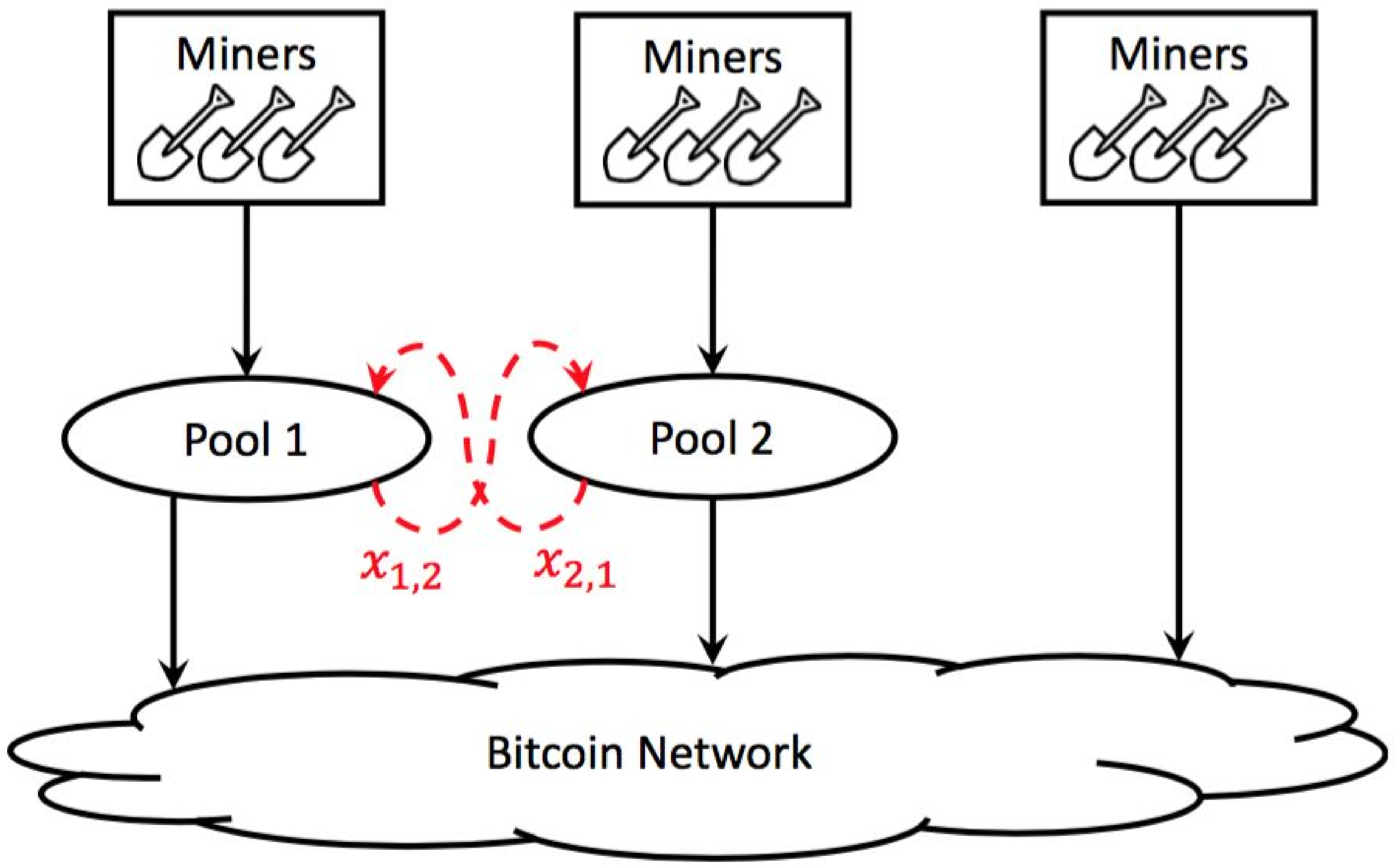


Fig. 7. Two pools attacking each other.

POOL WARS

TRAGEDY OF THE COMMONS

- No-pool-attacks is not a Nash equilibrium
  - If none of the other pools attack, a pool can increase its revenue by attacking the others
- But if the pools agree not to attack, both (or all) benefit in the long run.
  - However, this is an unstable situation since on a practical level you can attack another pool anonymously
- If pools can detect attacks then maybe an optimistic long term solution is feasible

Pool 2 \ Pool 1	no attack	attack
no attack	$(r_1 = 1, r_2 = 1)$	$(r_1 > 1, r_2 = \tilde{r}_2 < 1)$
attack	$(r_1 = \tilde{r}_1 < 1, r_2 > 1)$	$(\tilde{r}_1 < r_1 < 1, \tilde{r}_2 < r_2 < 1)$

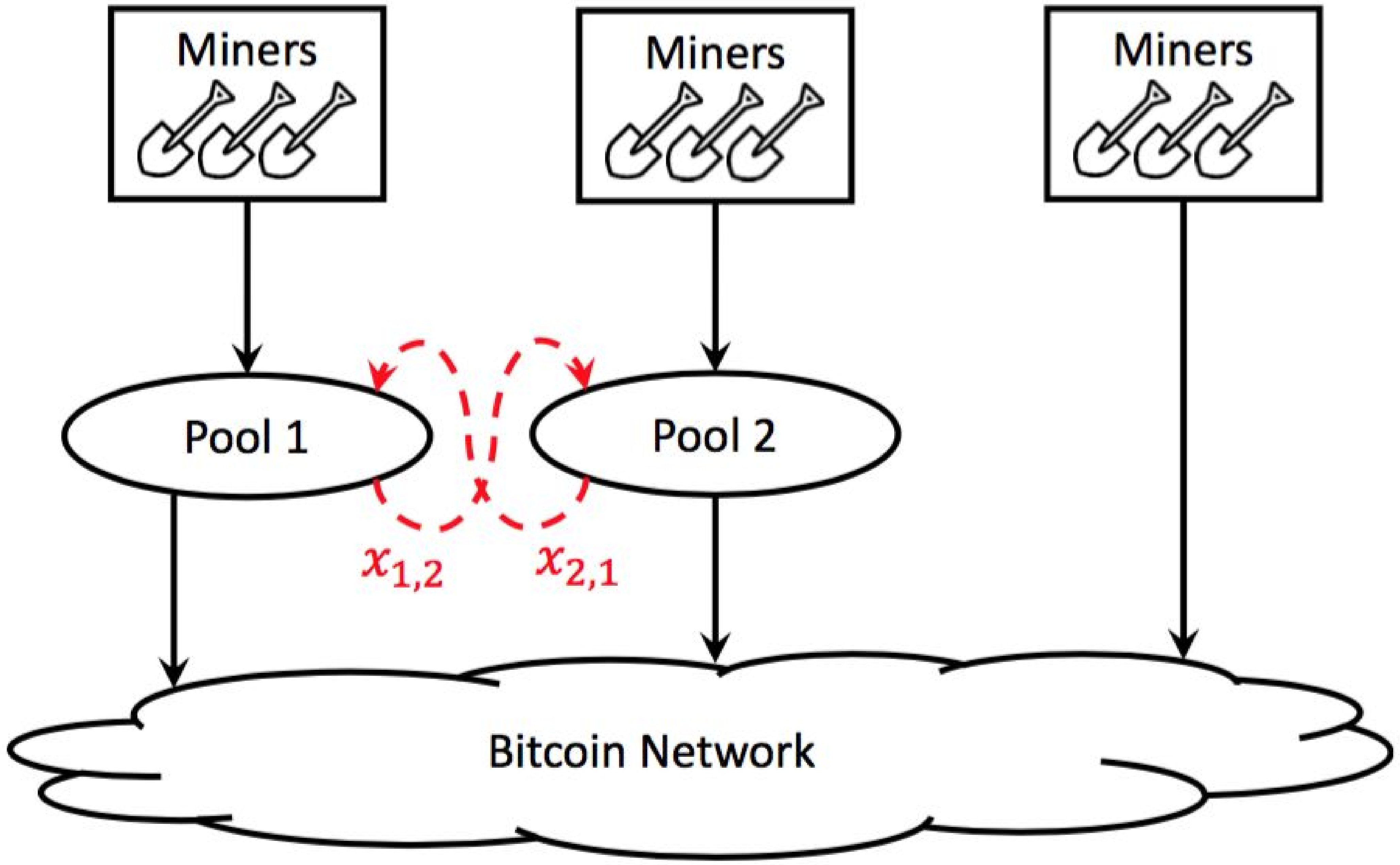


Fig. 7. Two pools attacking each other.

Nash Equilibrium is a Tragedy of the Commons

2 FORKING & DOUBLE SPENDS

DOUBLE SPENDING

THE CLASSIC ATTACK

Double Spend: Successfully spending the same money more than once.

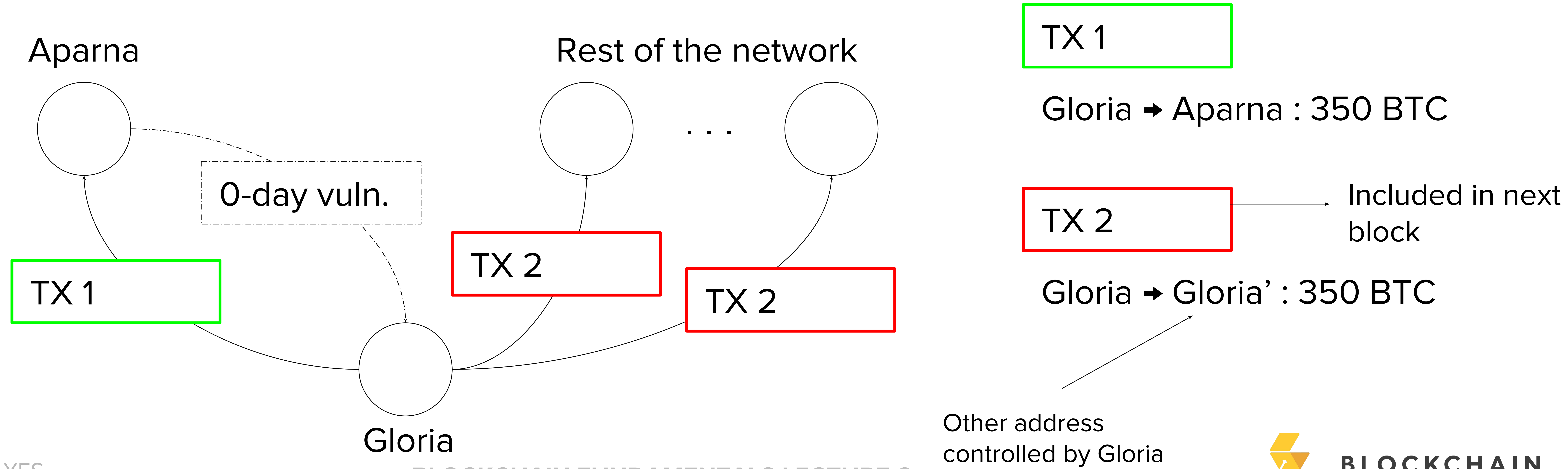
- Gloria wants to buy an iPhone 0-day exploit from Aparna on the black market for $1.5 million ~ 350 BTC but doesn't want to give up her bitcoins. #HODL
  - How can Gloria double spend on Aparna?

DOUBLE SPENDING

RACE ATTACK

Suppose Aparna simply checks that the transaction he sees is valid and **immediately** sends Gloria the iPhone.

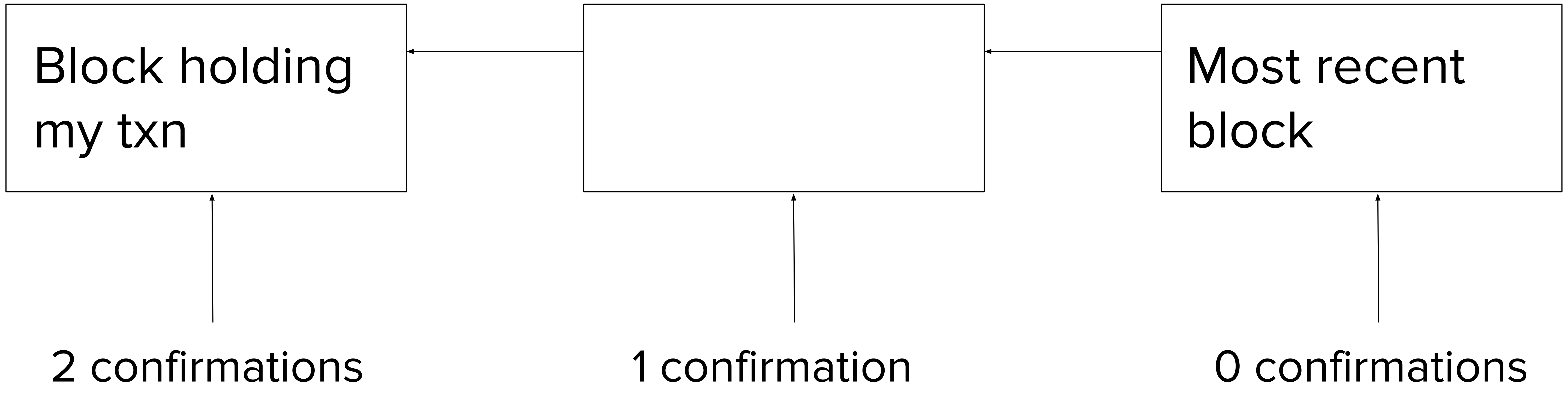
Aparna is vulnerable to a **Race Attack**!



DOUBLE SPENDING
CONFIRMATIONS

Confirmations: The number of blocks created on top of the block a txn is in.

```
confirmations = block_depth - 1
```



DOUBLE SPENDING

Z-CONFIRMATIONS

Clearly not secure if Aparna doesn't wait for any confirmations… What if Aparna waits for **z** confirmations?

[G → A] transaction needs **z** confirmations before Aparna sends the goods. In order to double spend on Aparna, Gloria needs to start a private chain containing her malicious transaction, mine **z** blocks on top of her block, then broadcast *after* Aparna sends the goods.

Now, Gloria has her money back, even though Aparna already sent the goods!

Suppose Aparna waits for **z** = 2 confirmations

Honest Network

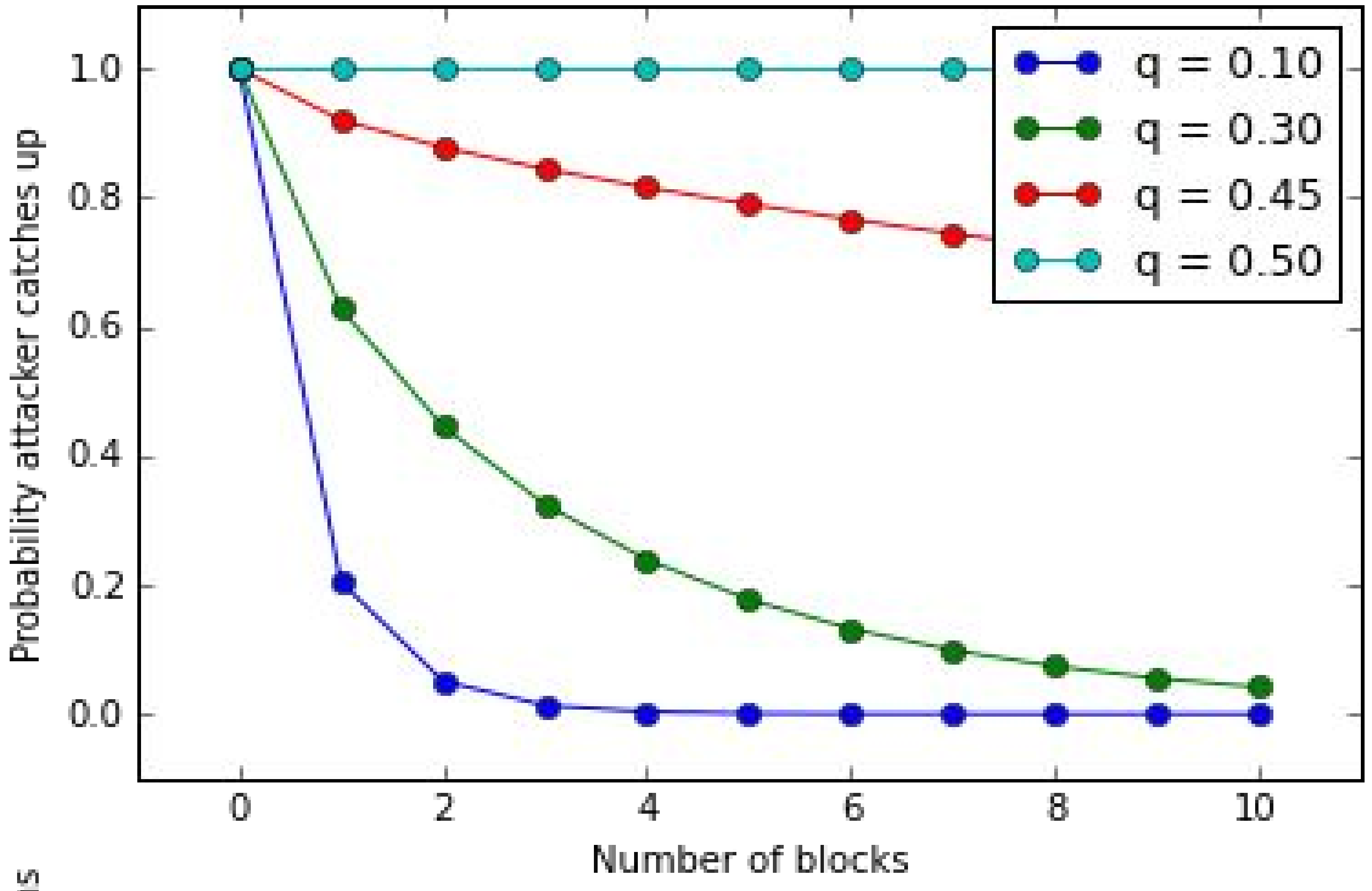
z = 2 confs, Aparna sends goods

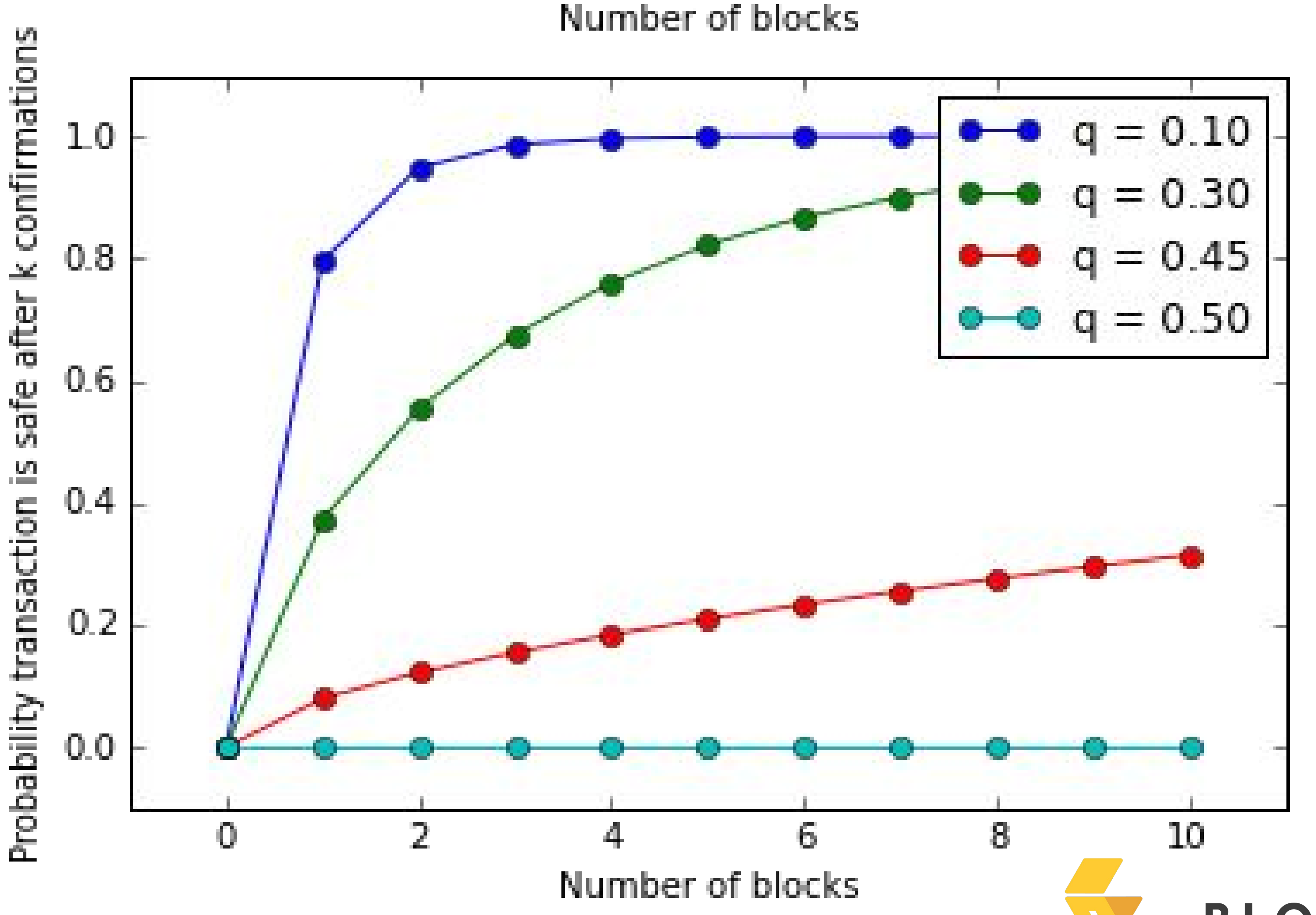
Gloria broadcasts her chain, which has a higher block height, so the honest network accepts Gloria's chain over the previous chain.

Gloria mining blocks, withholding from honest network

DOUBLE SPENDING

PROBABILITIES





Probabilities of success for the attacker

(Inverse represents bounds of the probability of safety for the vendor given assumptions of the attacker's hashpower)

AUTHOR: PHILIP HAYES

DOUBLE SPENDING

51% ATTACK

What if Gloria controls more than 50% of the total network hash power?

Whenever Gloria is **j** blocks behind the honest network, she will *always* (in expectation) be able to catch up and out-produce the honest miners.

Therefore, the probability that Gloria can successfully **double spend** with >50% hash power reaches 1!

DOUBLE SPENDING

INCENTIVES

Why would Gloria not want to double spend?

If the rest of the network detects the double spend, it is assumed that confidence in the cryptocurrency and exchange rate would *plummet*.

If Gloria isn't staked in Bitcoin she can *short* the currency to profit after her attempted double spend.

Bribing Miners:

Gloria might not physically control the mining hardware necessary to perform a double spend.

Instead, Gloria can bribe miners or even entire pools to mine on her withheld chain.

What if Gloria is a **hostile government** / **adversarial altcoin** / **large finance institution** with *significant capital* available?

Gloria can acquire enough mining ASICs or bribe enough miners / pools to achieve >50% effective hash power.

Gloria can perform a so-called **"Gold Finger" attack** with the objective of *destroying* the target cryptocurrency, either by destroying confidence in the currency with a double spend or spamming the network with empty blocks.

Ex: Eligius pool kills CoiledCoin altcoin [3]

3 CENSORSHIP

BLACKLISTING

FREEZE PEACH IS DEAD

You are a government that has jurisdiction over mining pools, say China (controlled entirely by Gloria).

Objective: Censor the Bitcoin addresses owned by certain people, say Gary Johnson, and prevent them from spending any of their Bitcoin

Block containing transactions from Gary Johnson

Normal block

Block mined by Chinese miners

BLACKLISTING

NAIVE CENSORSHIP STRATEGY

First strategy:

Tell Gloria’s mining pools not to include Johnson's transactions **(blacklisting)**

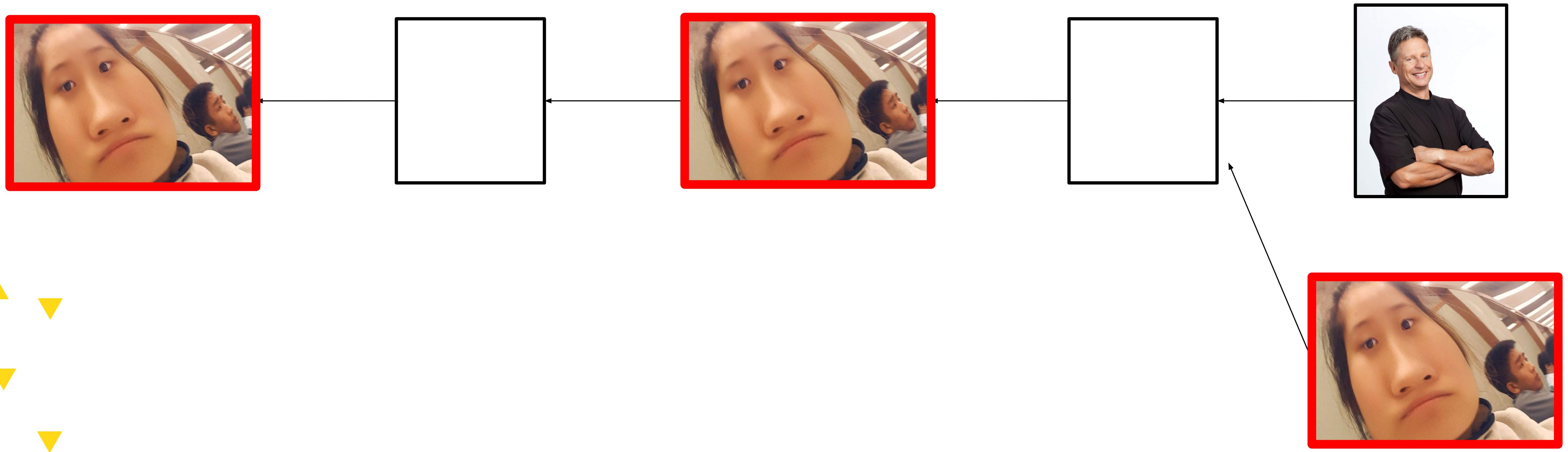
- Doesn't work unless you are 100% of the network
- Other miners will eventually include Gary’s transactions in a block
- Can only cause delays and inconveniences

BLACKLISTING

PUNITIVE FORKING

Second strategy:

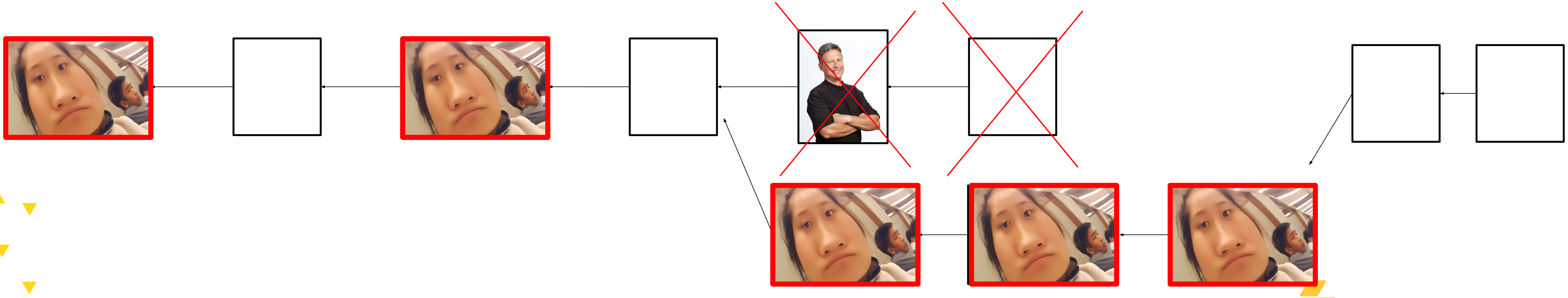
- Remember Gloria, you are China: **you have >51% of the network hashrate**
- Mandate that Chinese pools will refuse to work on a chain containing transactions spending from Gary's address
- Announce this to the world



BLACKLISTING

PUNITIVE FORKING

- If miners include a transaction from Gary in a block, Gloria will fork and create a longer proof-of-work chain
- Block containing Gary's transaction now invalidated, can never be published

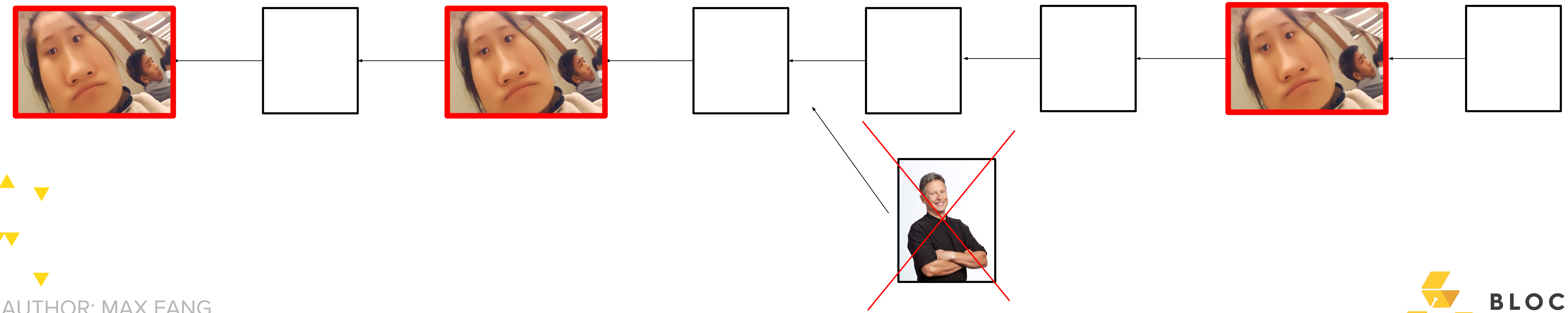


BLACKLISTING

PUNITIVE FORKING

- Non-Chinese miners eventually stop trying to include Gary's transactions when mining blocks, since they know that their block will be invalidated by Chinese miners when they do

We have now shown how a 51% majority can prevent anyone from accessing their funds. This is called **punitive forking.**



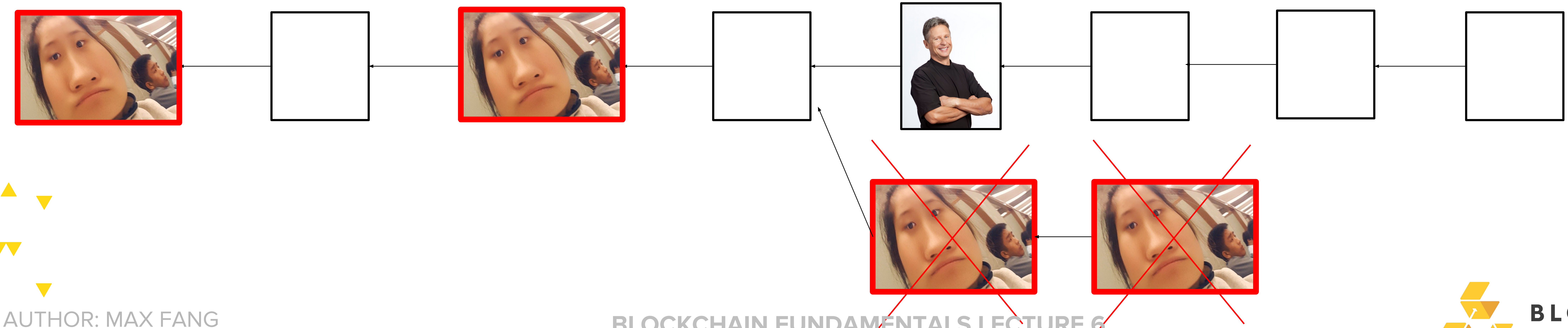
AUTHOR: MAX FANG

BLACKLISTING

FEATHER FORKING

Punitive forking doesn't work unless you have >51% of hashpower. Is there another way?
Yes! Called **Feather Forking**

- New strategy: Announce that you will **attempt** to fork if you see a block from Gary, but you will give up after a while
  - As opposed to attempting to fork forever; doesn't work without >51%
- Ex. Give up after block with Gary's tx contains **k** confirmations



BLACKLISTING

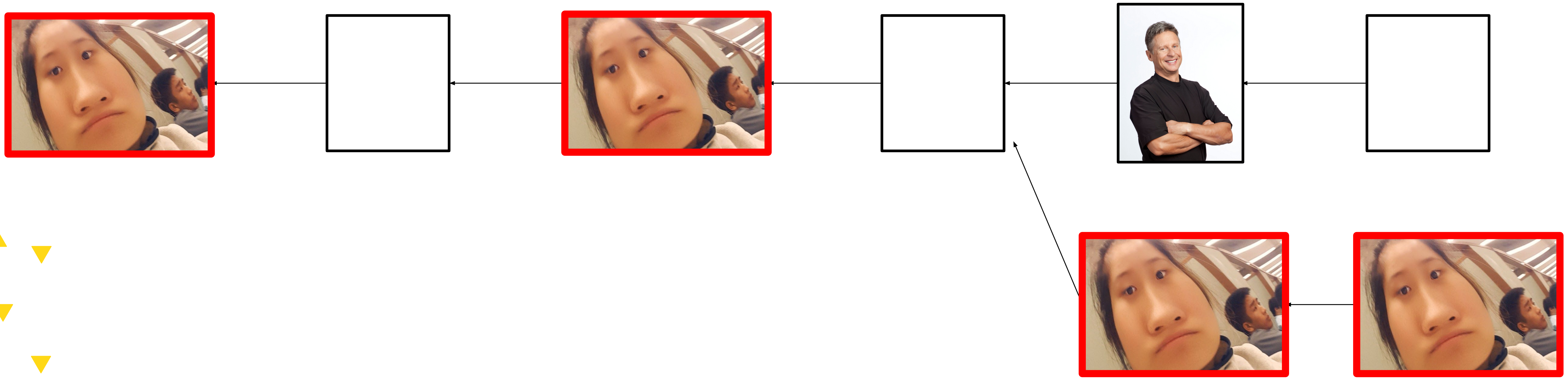
FEATHER FORKING

Let **q** equal the proportion of mining power you have, $0 < \mathbf{q} < 1$

Let **k** = 1: You will give up after 1 confirmation (one additional block)

- Chance of successfully orphaning (invalidating) the Johnson block = $\mathbf{q^2}$

If **q = .2**, then $\mathbf{q^2}$ **= 4%** chance of orphaning block. Not very good



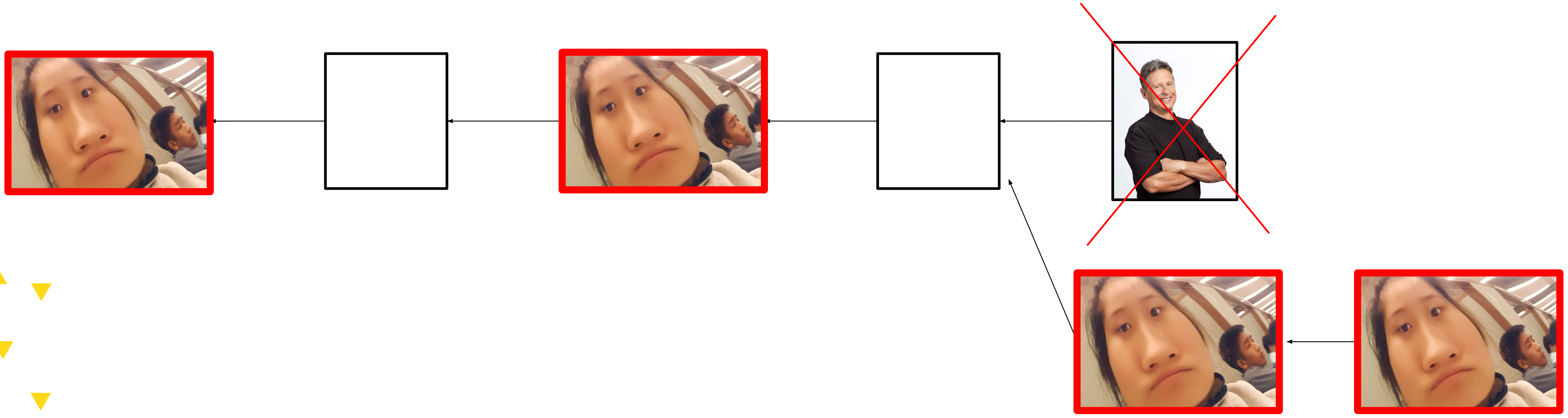
BLACKLISTING

FEATHER FORKING

But other miners are now aware that their block has a $\mathbf{q^2}$ chance of being orphaned. They must now decide whether they should include Johnson's tx in their block

EV(include) = (1 - $\mathbf{q^2}$) * BlockReward + Johnson's tx fee

EV(don't include) = BlockReward



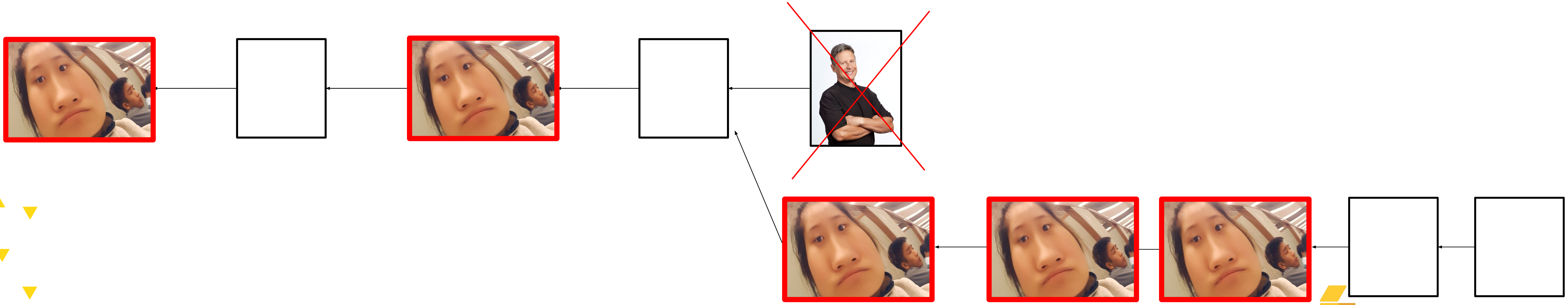
BLACKLISTING

FEATHER FORKING

EV(include) = (1 - **q^2**) * BlockReward + Johnson's tx fee
EV(don't include) = BlockReward

Therefore, unless Gary Johnson pays **q^2** * BlockReward in fees for his transaction, other miners will mine on the malicious chain

- 4% * 12.5 BTC = 0.5 BTC = Johnson must pay **$2409.50** minimum/transaction



4

SELFISH MINING

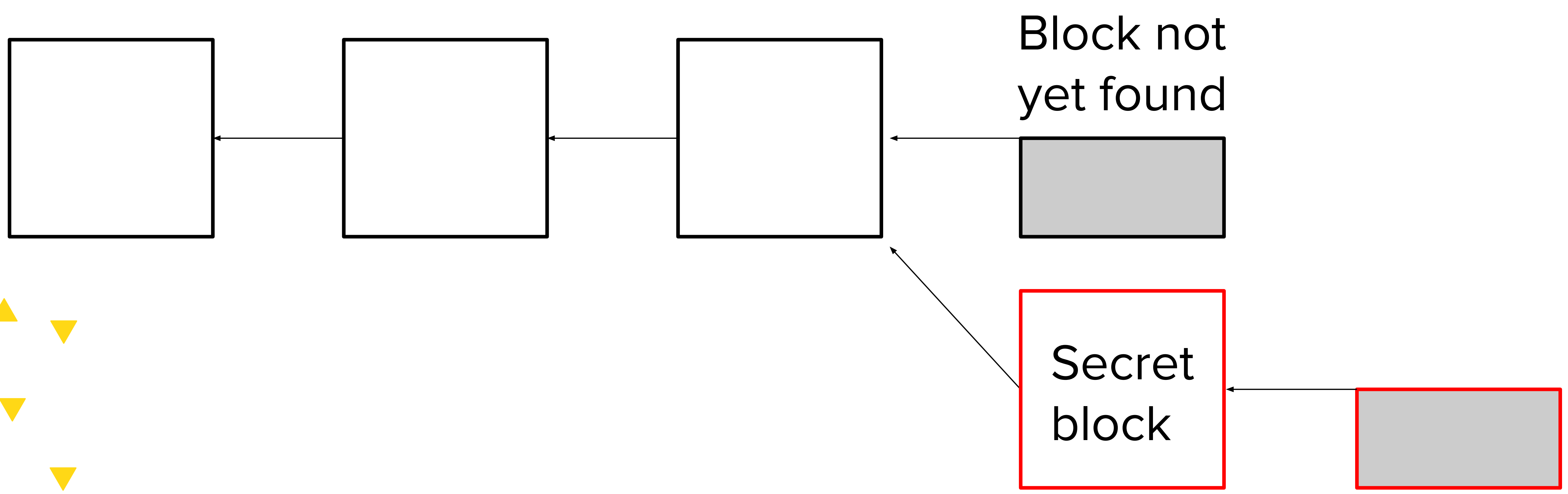
SELFISH MINING

BLOCK WITHHOLDING

You are a miner; suppose you have just found a block.

- Instead of announcing block to the network and receiving reward, keep it secret
- Try to find two blocks in a row before the network finds the next one

This is called **selfish mining** or **block-withholding**



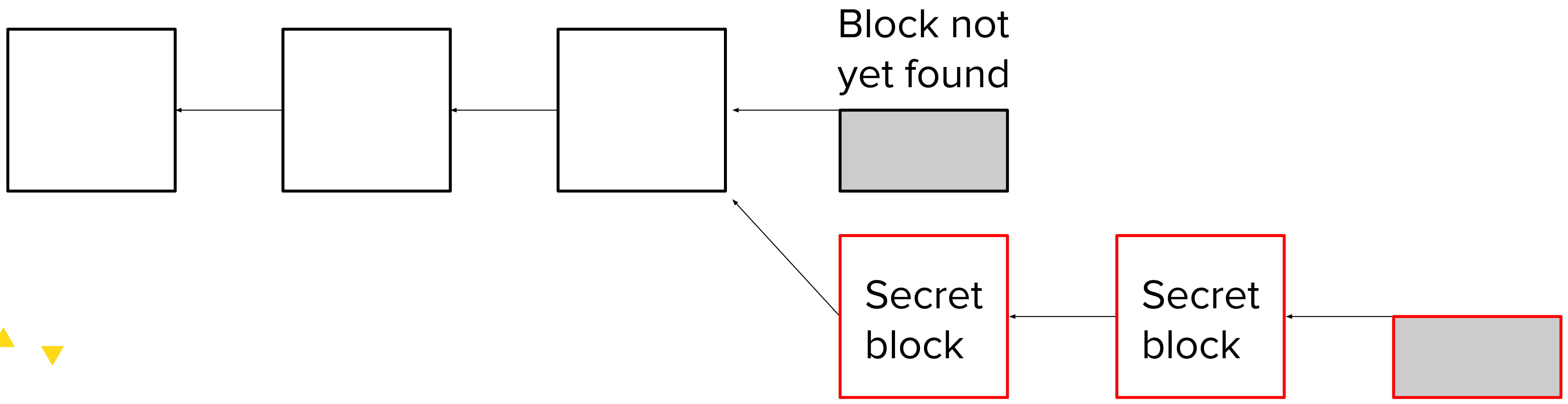
Note: "block-withholding" is also sometimes used in the context of mining pools - submitting shares but withholding valid blocks

SELFISH MINING

BLOCK WITHHOLDING

If you succeed in finding a second block, you have fooled the network

- Network still believes it is mining on the longest proof of work chain
- You continue to mine on your own chain

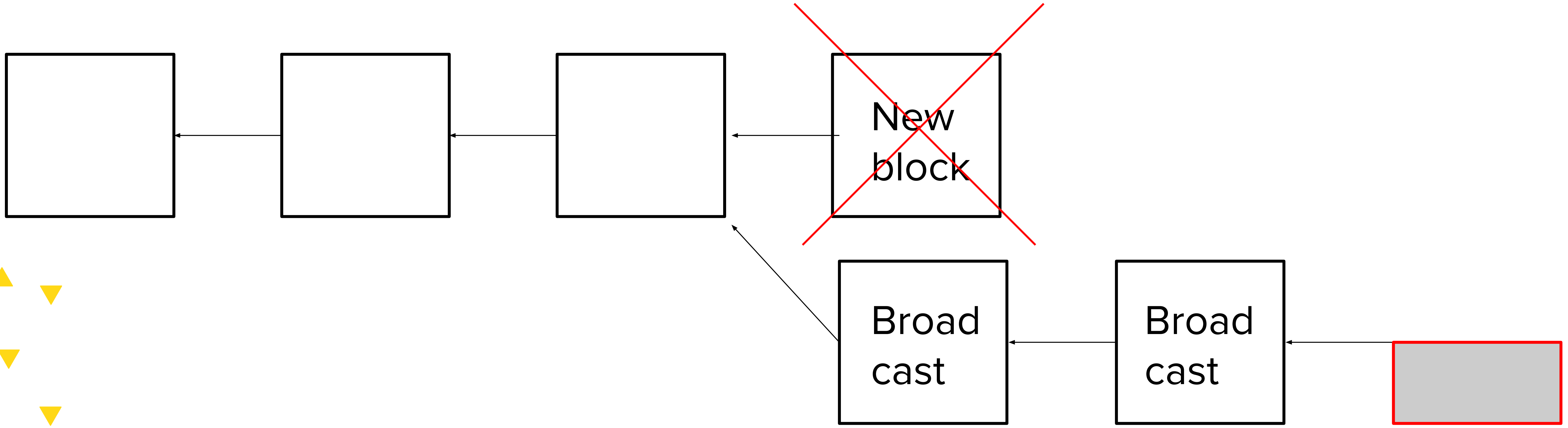


SELFISH MINING

BLOCK WITHHOLDING

If the network finds a block, you broadcast your two secret blocks and make the network block invalid

- While network was working on the invalid block, you got a bunch of time to mine by yourself… for free!
- Free time mining on network
  => higher effective proportion of hashrate **=> higher expected profits!**

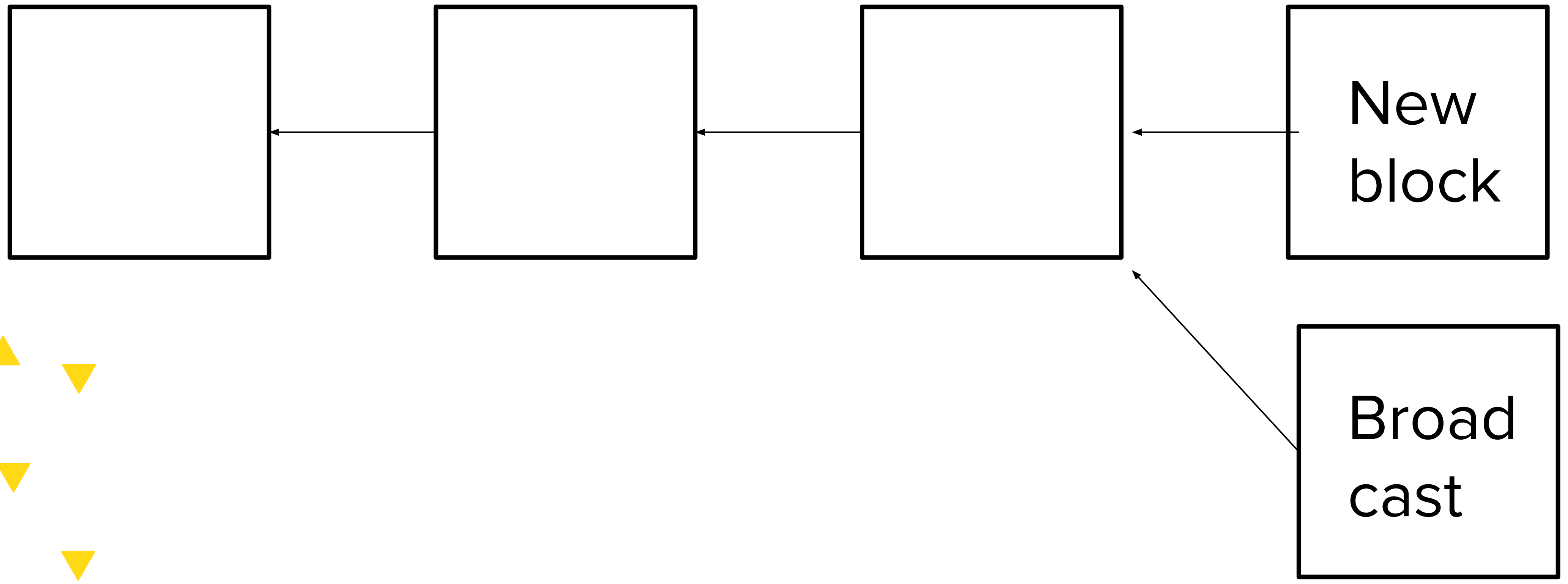


SELFISH MINING

BLOCK WITHHOLDING

But what if the network found their new block before you could find a second one? **Race to propagate!**

- If on average you manage to tell 50% of the network about your block first:
  - Malicious strategy is more profitable if you have >25% mining power
- If you have >33% mining power, **you can lose the race every time and malicious strategy is still more profitable!**
  - (actual math omitted due to complexity)



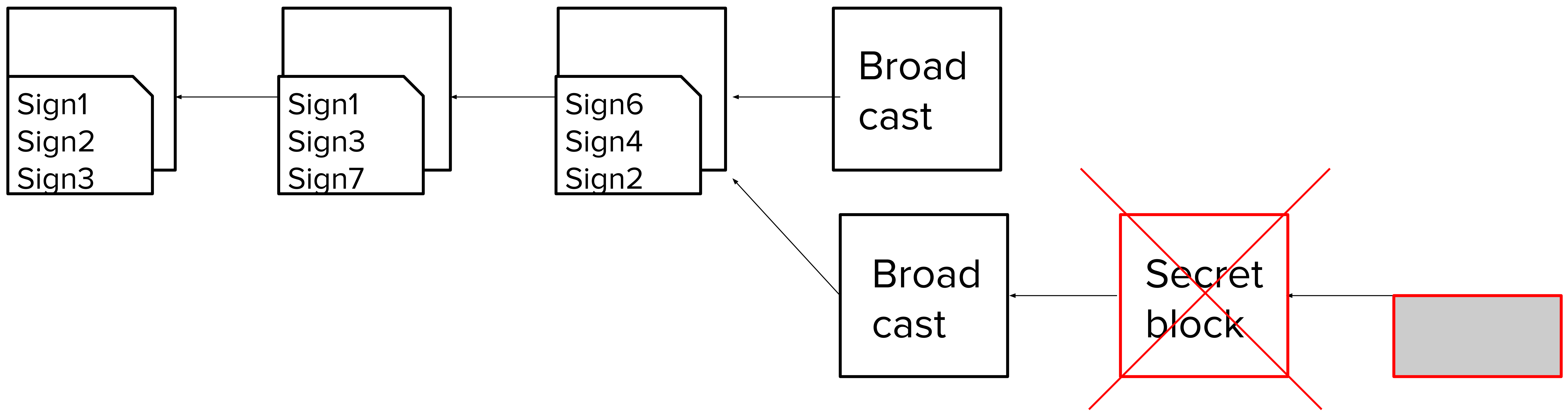
5 PUBLISH OR PERISH: A DEFENSE

DEFENSES: BLOCK VALIDATION

DUMMY BLOCK SIGNATURES

Proposed by Schultz (2015), Solat and Potop-Butucaru (2016)

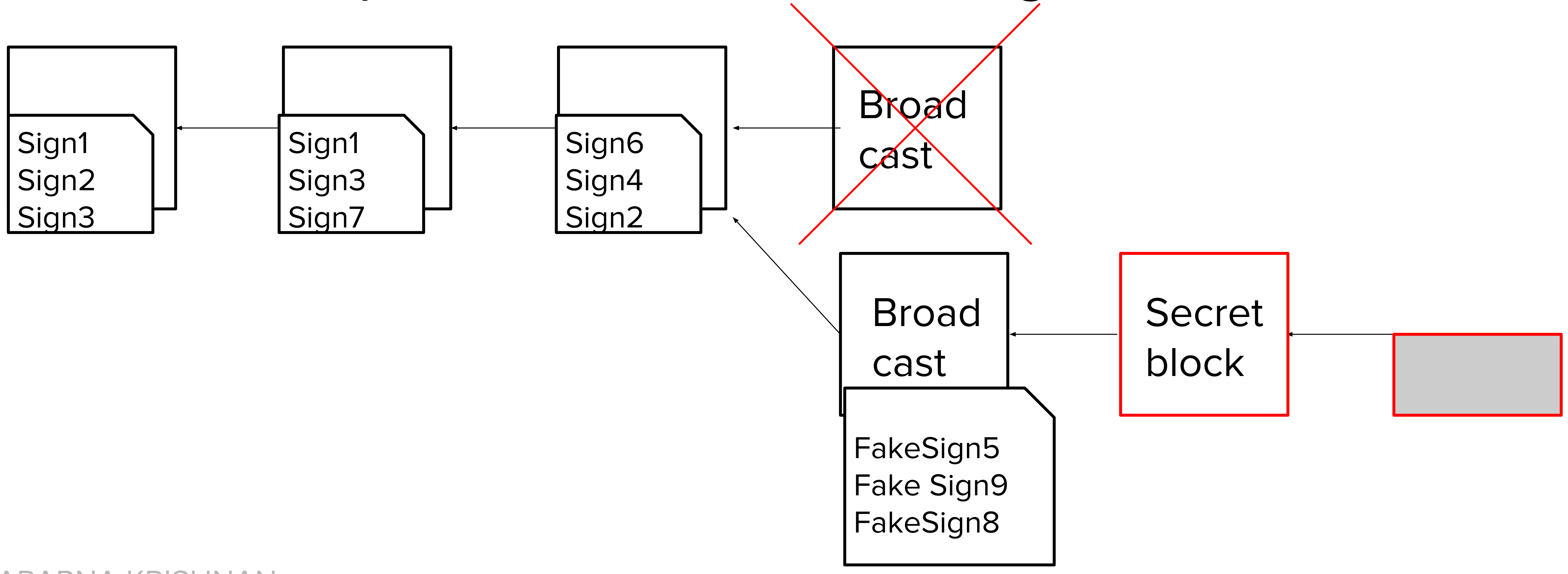
- Accompany solved blocks with signatures on dummy blocks
- Proves that the block is witnessed by the network
  - Proves that a competing block is absent before miners are able to work on it



DEFENSES: BLOCK VALIDATION

DUMMY BLOCK SIGNATURES

- However, does not provide a mechanism to evaluate whether the number of proofs is adequate to continue working
- Does not discuss how to prevent Sybil attacks on signatures
  - Selfish miner generates many signatures on the dummy block
- This defense requires fundamental changes to the block validity rules



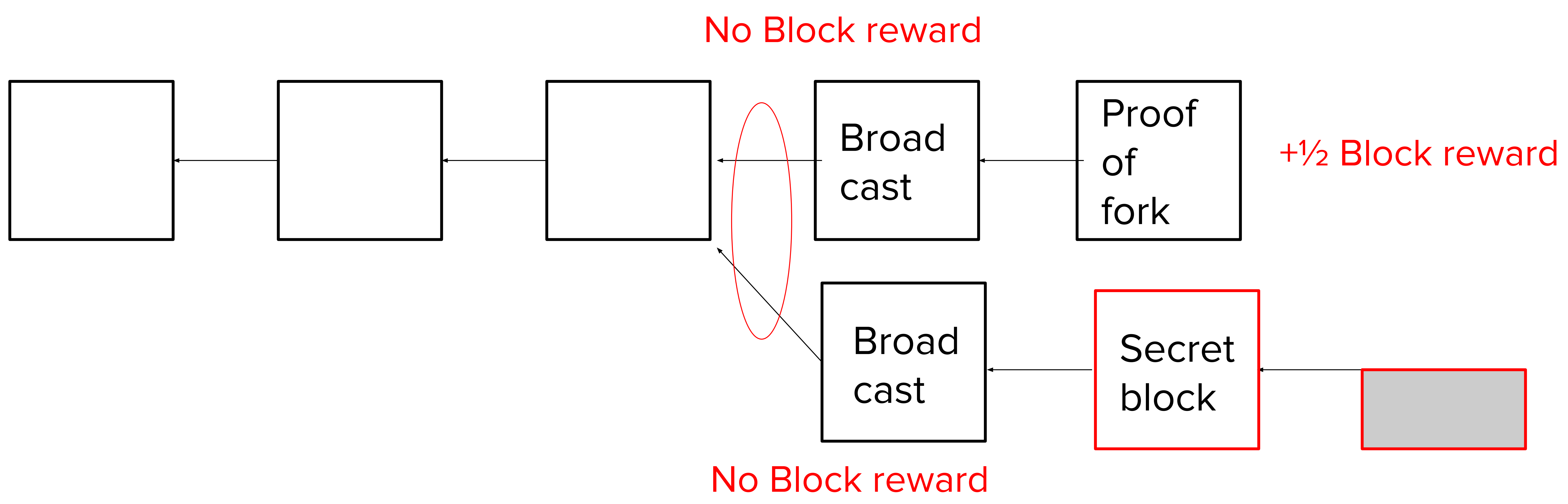
BLOCKCHAIN AT BERKELEY

DEFENSES: FORK-PUNISHMENT

FORK-PUNISHMENT RULE

Proposed by Lear Bahack (2013)

- Competing blocks receive no block reward
- The first miner who incorporates a poof of the block fork in the blockchain gets half of the forfeited rewards



DEFENSES: FORK-PUNISHMENT

FORK-PUNISHMENT RULE

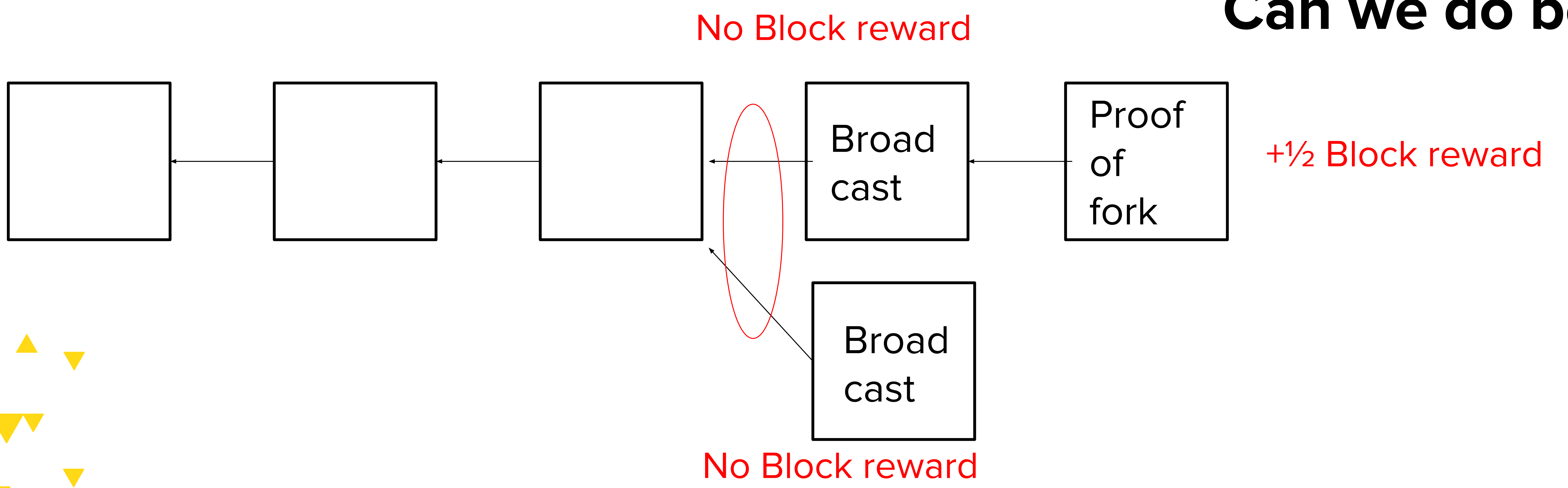
Drawbacks

- Honest miners suffer collateral damage of this defense
  - This defense constitutes another kind of attack

This defense requires fundamental changes to the reward distribution rules

- Requires a hard fork to implement
  - We have hard enough time fixing transaction malleability

Can we do better?

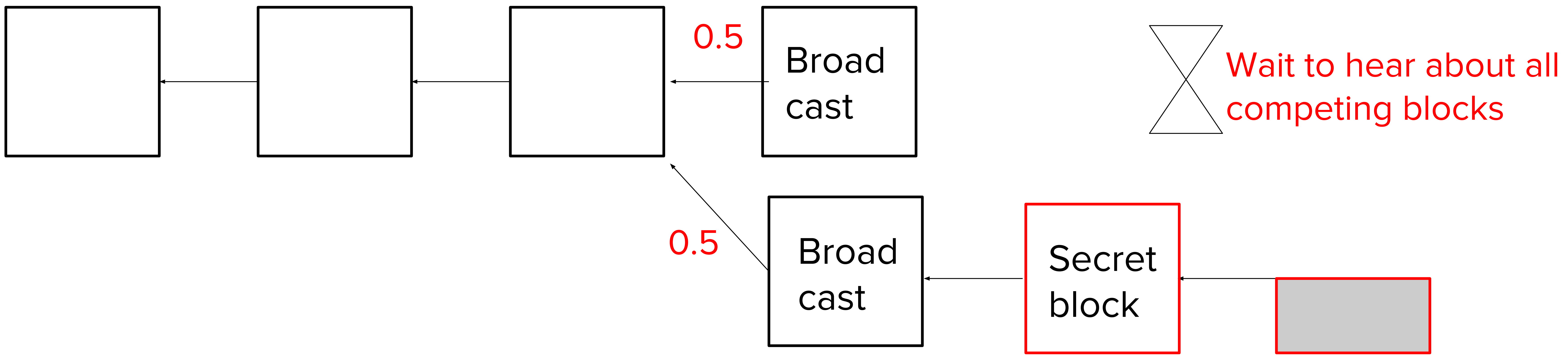


DEFENSES: TIE-BREAKING

UNIFORM TIE BREAKING

Proposed by Eyal and Sirer (2014)

- In the case of a tie, a miner randomly chooses which chain to mine on
  - Prevents an attacker from benefiting from network-level dominance
- Raises the profit threshold from 0% to **25%** under their strategy
  - Sapirshtein (2015) proposes a more optimal selfish mining strategy
  - Reduces Eyal and Sirer profit threshold to **23.2%**



AUTHOR: MAX FANG, APARNA KRISHNAN

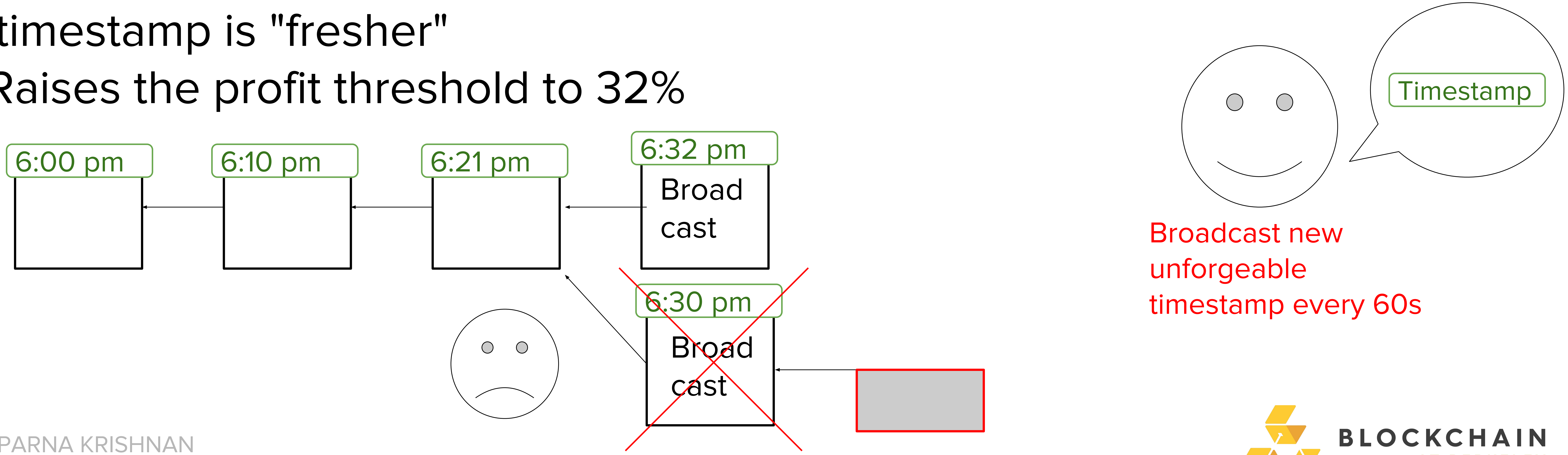
BLOCKCHAIN AT BERKELEY

DEFENSES: TIE-BREAKING

UNFORGEABLE TIMESTAMPS

Proposed by Ethan Heilman (2014)

- Each miner incorporates the latest unforgeable timestamp issued by a trusted party into the working block
  - Timestamp is publically accessible and unpredictable
  - Issued with an interval of 60s
- When two competing blocks are received within 120s, a miner prefers the block whose timestamp is "fresher"
- Claim: Raises the profit threshold to 32%

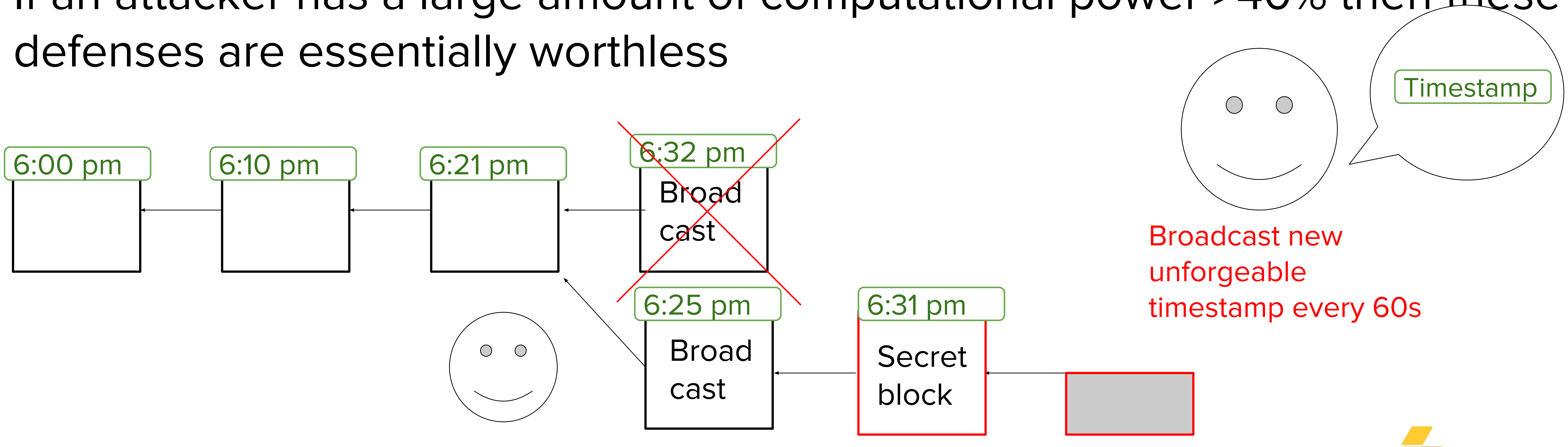


DEFENSES: TIE-BREAKING

UNFORGEABLE TIMESTAMPS

Drawbacks

- Tie-breaking rules don't apply when the selfish mining chain is longer than the public chain
  - Only applies to a block propagation race
- If an attacker has a large amount of computational power >40% then these defenses are essentially worthless



DEFENSES: TIE-BREAKING

UNFORGEABLE TIMESTAMPS

Drawbacks

- Introducing central party contradicts the bitcoin philosophy

+ CENTRALISATION

PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Ren Zhang and Bart Preneel (Apr 2017) claim the best-yet defense of selfish mining

- Backwards compatible: No hard fork
- Disincentivizes selfish mining even when if the selfish miner has a longer chain

Approach: A novel **F**ork-**R**esolving **P**olicy (FRP)

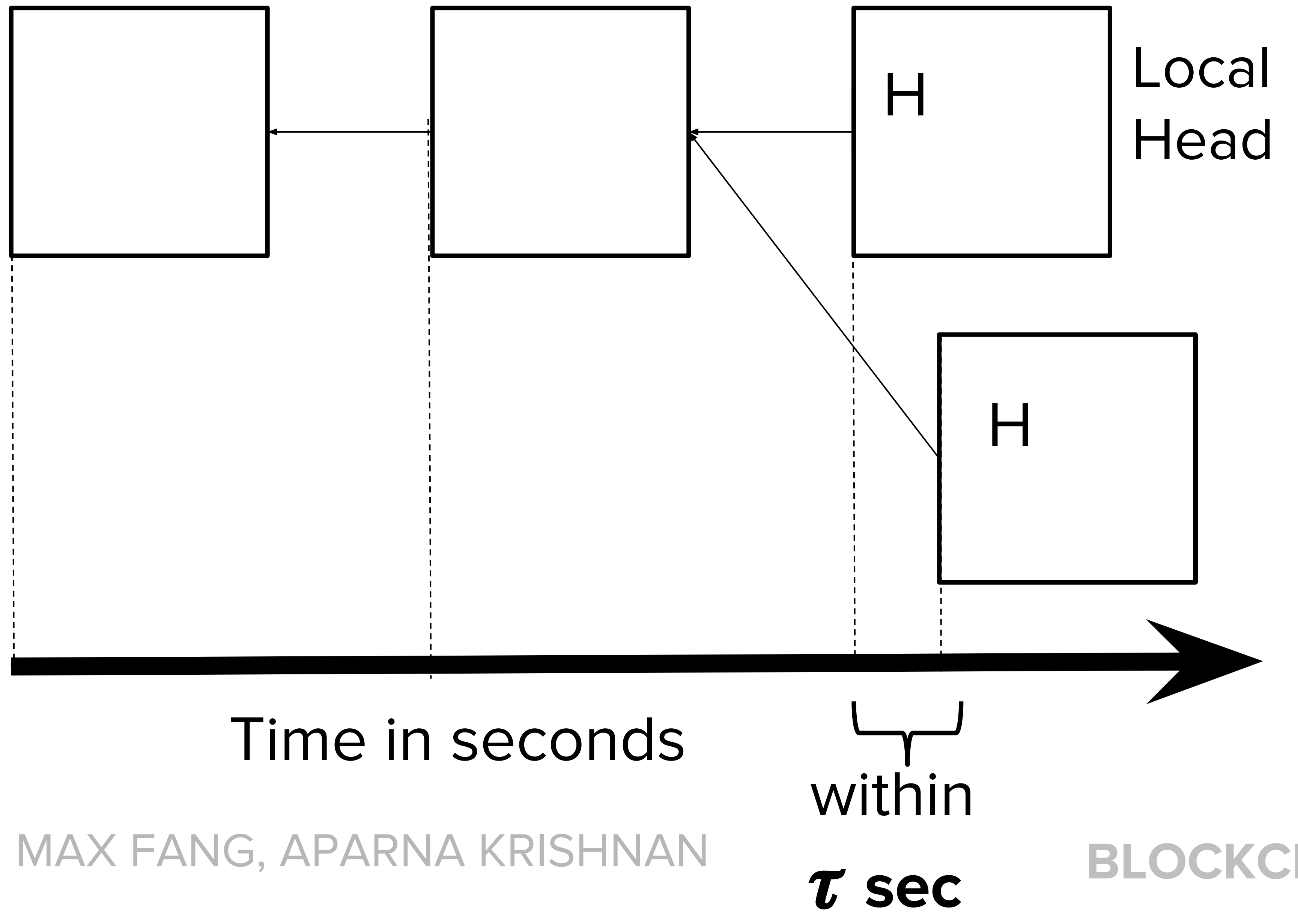
- Replace the original Bitcoin FRP (length FRP), with a **weighted FRP**
  - Embed in the working block the hashes of all its uncle blocks
- Note that selfish mining is premised on the idea of first building a secret block
- Idea: Make sure this secret block does not help the selfish miner win the block race

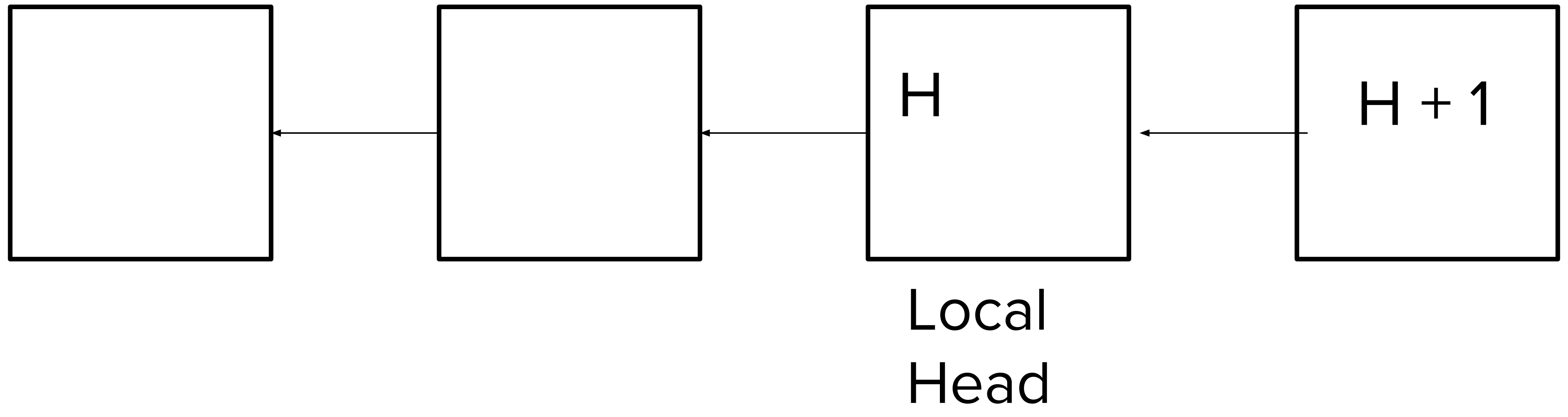
PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Definitions

- τ: An assumed upper bound on the amount of time it takes to propagate blocks across the Bitcoin network
- **In time.** Evaluated from the miner's local perspective.
  1. Height value is greater than that of the local head OR
  2. Height value is same as that of the local head, but was propagated within τ time



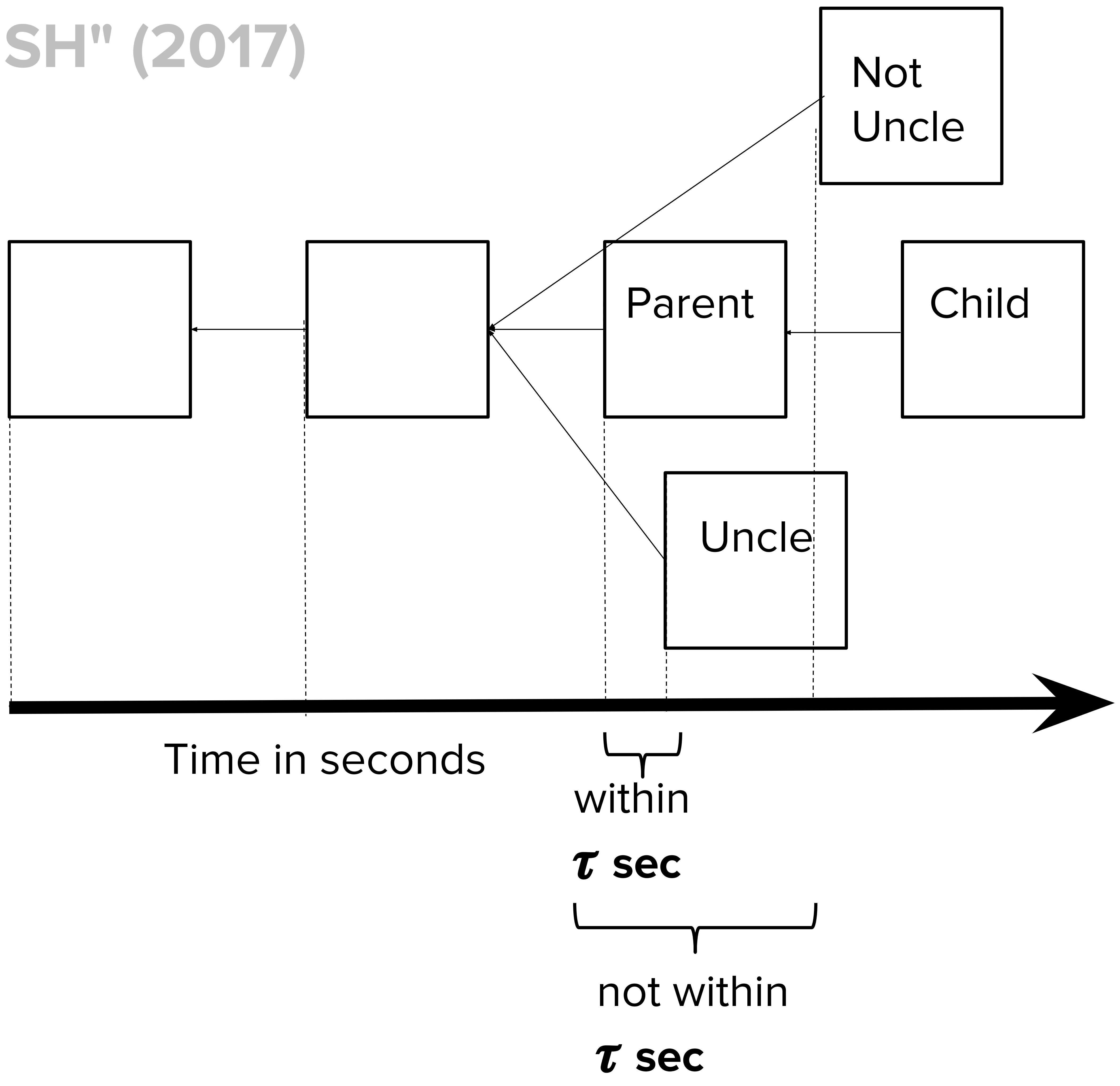


PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Definitions

- **Uncle**. Different from Ethereum's definition of an uncle
    1. The uncle of a block B is one less the height of B
    2. The uncle has to be in time
       *"A block B1 is the **uncle** of another block B2 if B1 is a competing in-time block of B2's parent block"*
- **Weight**. Since two competing chains always have a shared root, only consider blocks after that
    - weight = # of in time blocks + # of uncle hashes embedded in these blocks



BLOCKCHAIN AT BERKELEY

PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Zhang and Preneel's **Weighted Fork Resolving Policy**:

1. If one chain is longer *height-wise* than the other(s) by **k** or greater blocks*
   a. The miner will mine on this chain
2. Otherwise, the miner will choose the chain with the largest *weight*
3. If the largest weight is achieved by multiple chains simultaneously, then the miner chooses one among them randomly

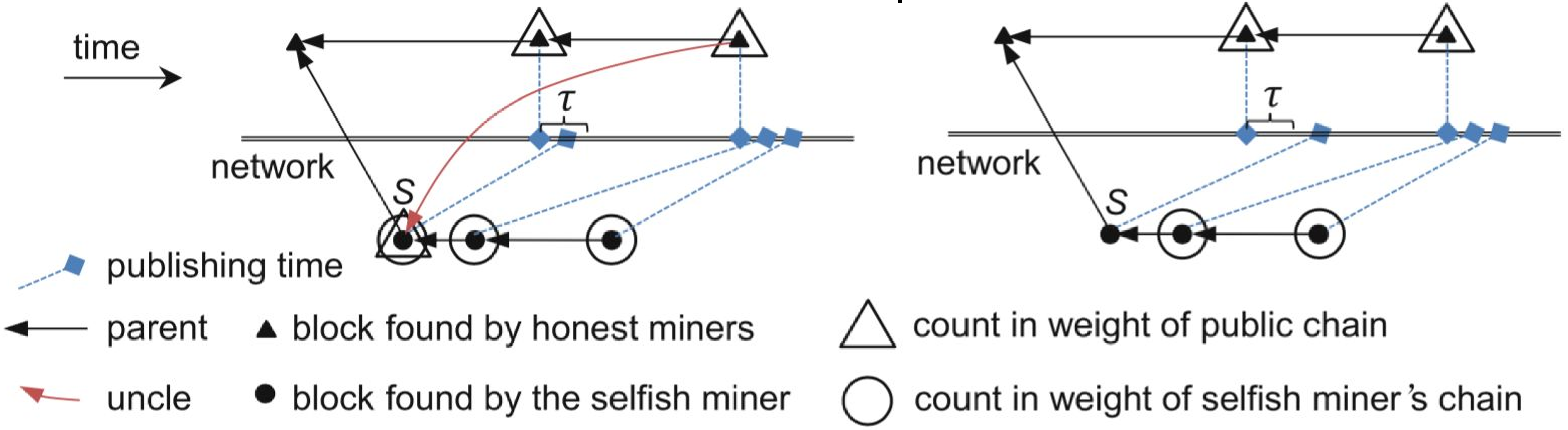
Aside:* ***k* *is a "fail-safe parameter" that gauges the allowed amount of network partition. Note that when* ***k*** *= ∞ the first rule never applies.*

PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Miner has one secret block. A competing block is published. **Block race**! Miner has two options:

- Option 1: If the selfish miner **publishes** their block, *the next honest block gains a higher weight* by embedding a proof of having seen this block
- Option 2: If the selfish miner **keeps their block secret**, the secret block *does not contribute* to the weight of its own chain
- In both scenarios, the secret block does not help the selfish miner win the block race



Choice 1: Publish **Choice 2: Don't publish**

PUBLISH OR PERISH

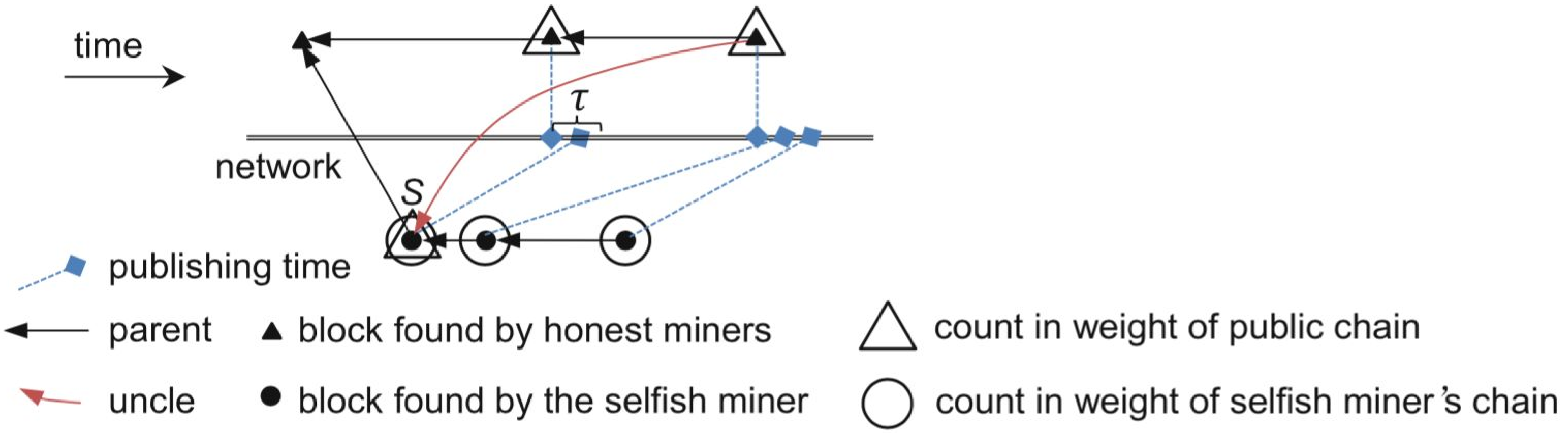
ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Same scenario revisited. More rigorously, let S be the first selfish block

Choice 1: Selfish miner publishes S

- S will be an uncle of the next honest block
  - (since it was published *in time* and its *height is one less*)

=> S counts into the weight of **both** the honest and the selfish chain



Choice 1: Publish

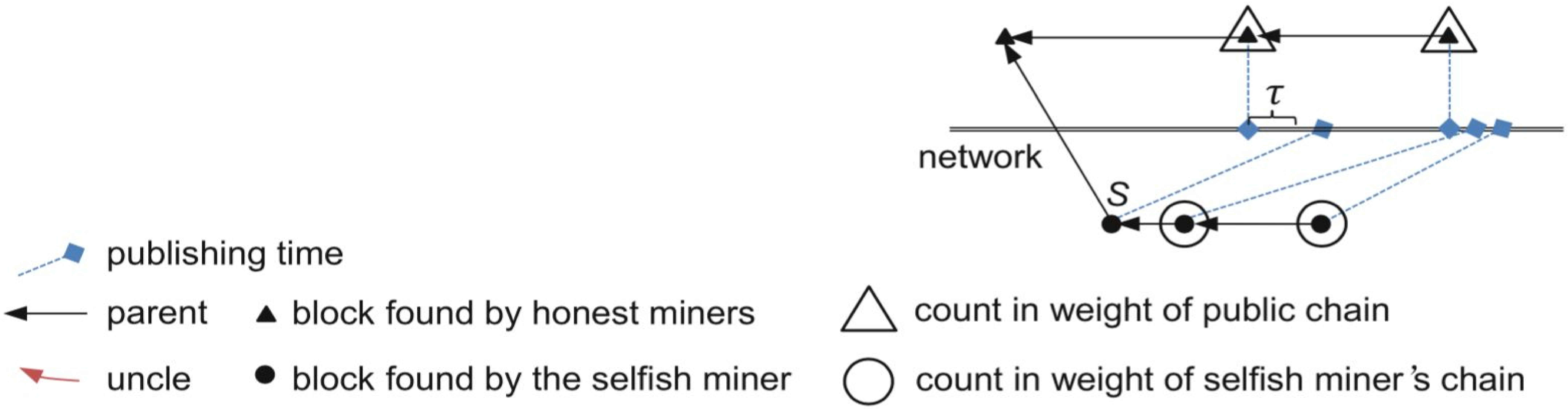
PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Choice 2: Selfish miner doesn't publish S

- Selfish miner waits, and publishes it later as a part of the selfish chain
- Honest miners do not count S into the weight of the selfish chain because S is not *in time*.
  - It is a **late block**
- S is not an *uncle* of the next honest block because the honest miners did not see it

=> S contributes to **neither** the weight of the honest nor the selfish chain



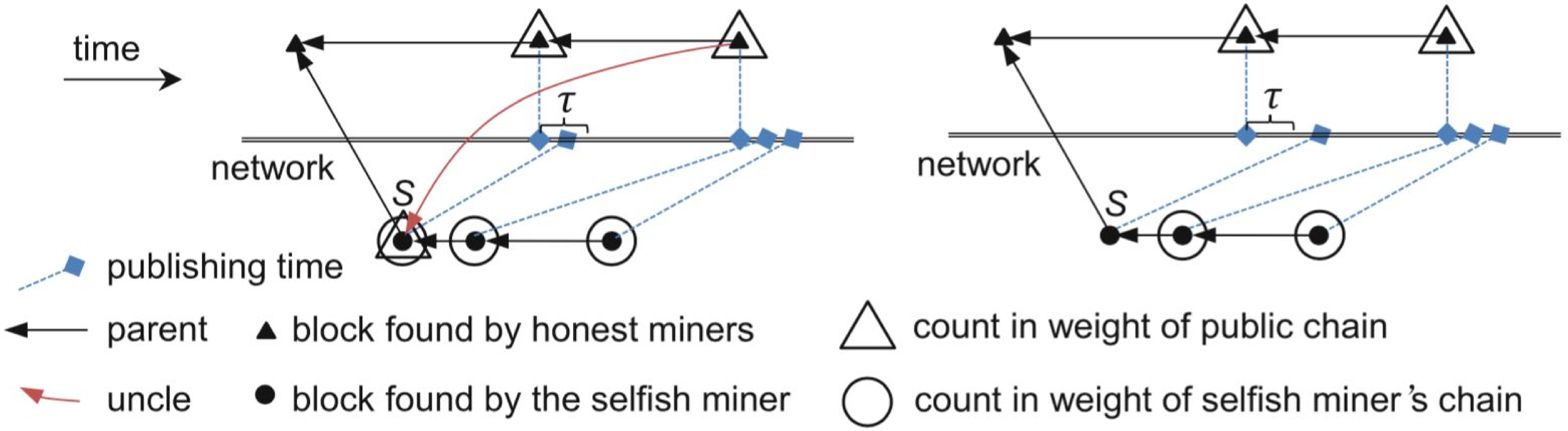
Choice 2: Don't publish

PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Result: Regardless of which option is chosen...

- S will **not** contribute to **only** the weight of the selfish chain.
    - Will only contribute to **both** or **neither**
- Completely nullifies the advantage of the secret block S !



Choice 1: Publish

Choice 2: Don't publish

PUBLISH OR PERISH

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

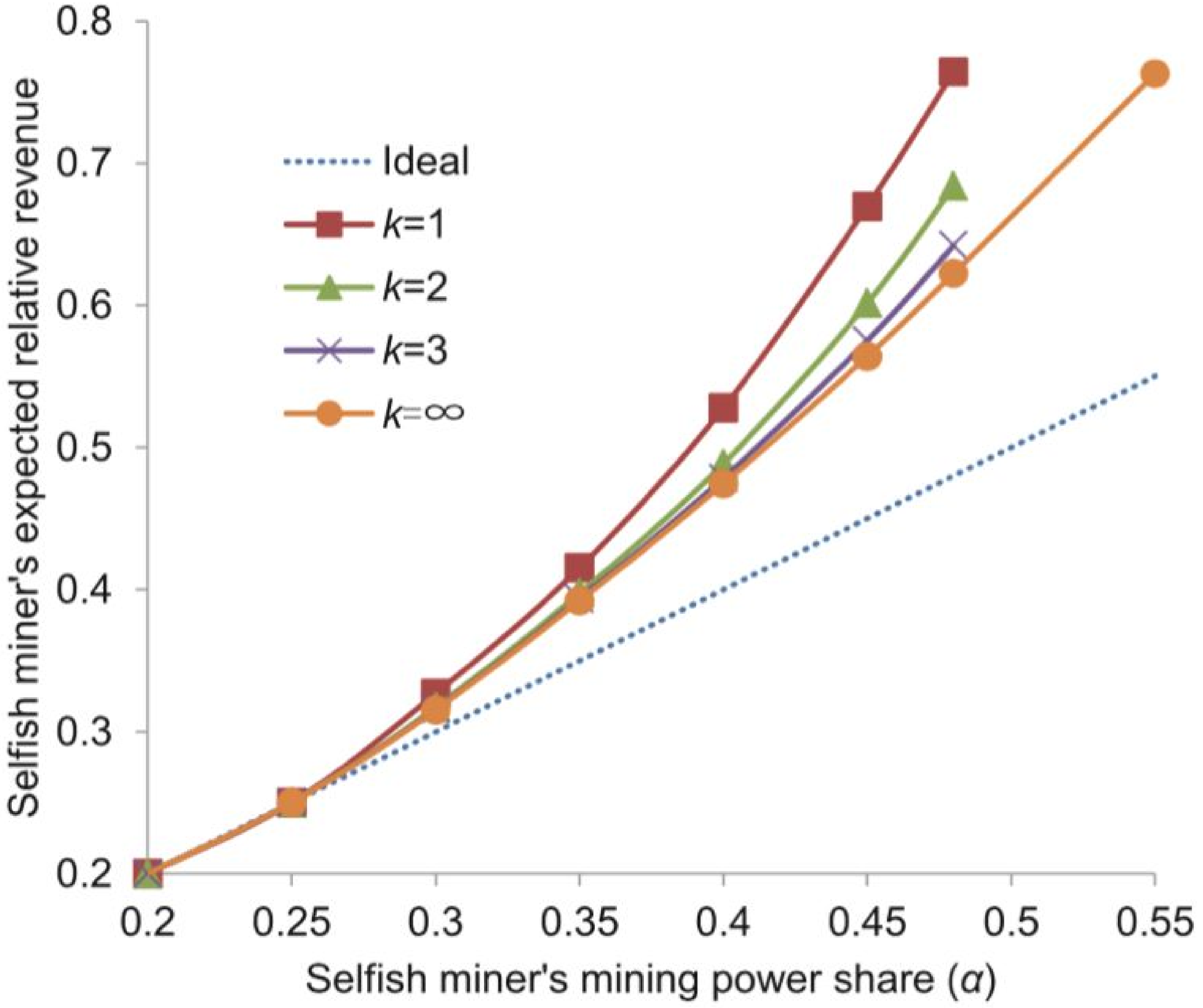


Fig. 4. Relative revenue of the selfish miner within our defense

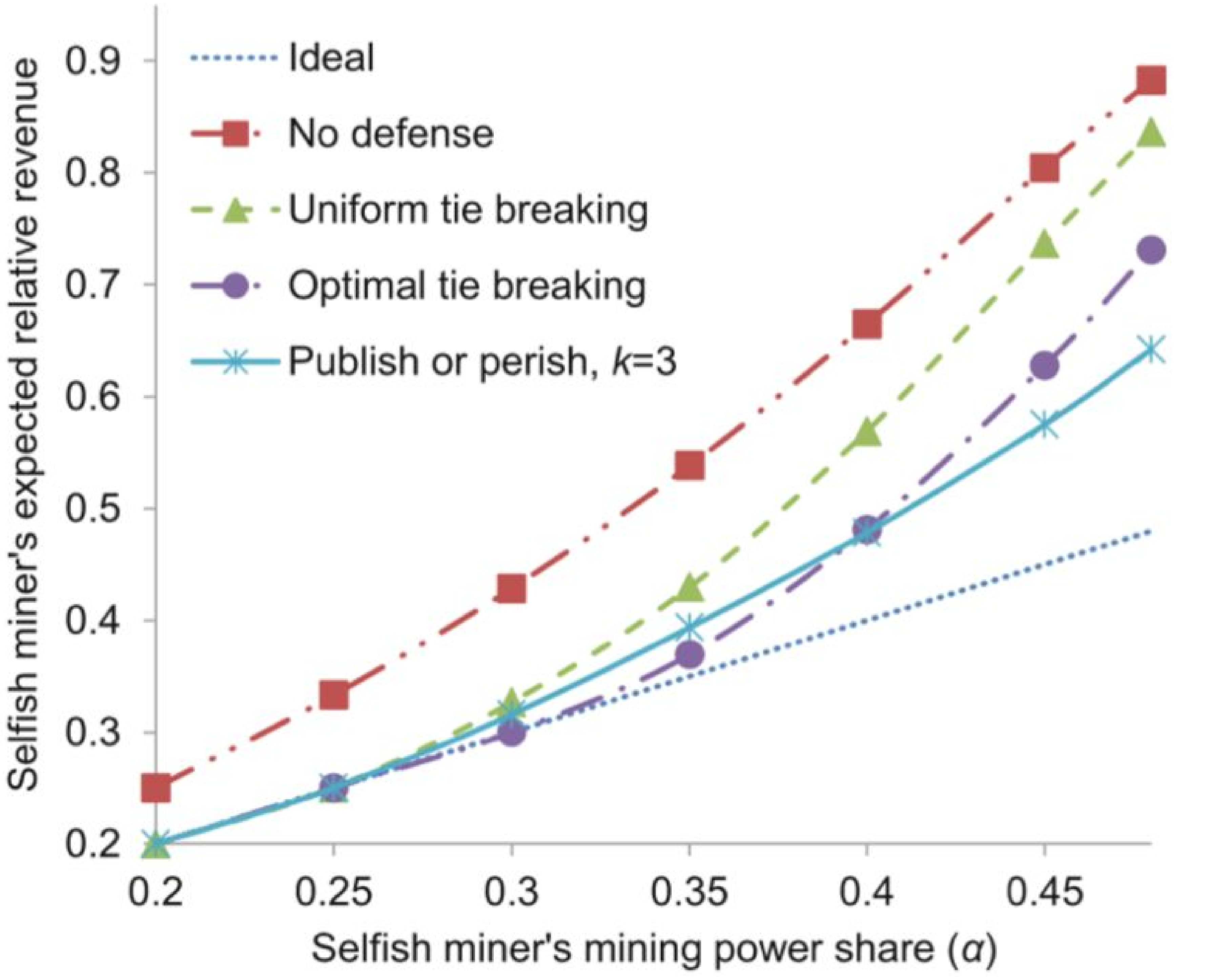


Fig. 5. Comparison with other defenses

PUBLISH OR PERISH: LIMITATIONS

ZHANG AND PRENEEL, "PUBLISH OR PERISH" (2017)

Limitations

- Bitcoin aims be asynchronous, Publish and Perish assumes synchronicity
  - (Because it assumes an upper bound of block propagation time)
  - Because of this, it's basically useless
- When the fail-safe parameter k > 1, an attacker may broadcast blocks right before they are late to cause inconsistent views among the honest miners
  - Several other selfish mining defenses also require a fixed upper bound on the block propagation time in order to be effective
- During the transition period to weighted FRP, an attacker can launch double-spend attacks
- Neglects real world factors:
  - Does not permit the occurrence of natural forks
  - Does not consider transaction fees on the selfish miner’s strategy
  - Does not consider how multiple selfish miners could collude and compete with each other
- Does not achieve incentive compatibility, but is the closest scheme to date

HOMEWORK

- HW:
  - Two options:
    - **Partner project**: One partner comes up with an attack, and the other comes up with a defense
    - ALTERNATIVE: Deep Dive! Alex is hosting an Alternative Consensus (with focus on Proof-of-Stake) Deep Dive **this Saturday** from 12 - 4 PM in HP Auditorium
      - Event: https://www.facebook.com/events/126620567999233
      - (Don’t need to stay the whole time)

READINGS

- Readings:
  - Bitcoin and Cryptocurrency Technologies - Princeton University
    - Ch 8.5 Proof-of-Stake and Virtual Mining
  - https://medium.com/@VitalikButerin/a-proof-of-stake-design-philosophy-506585978d51