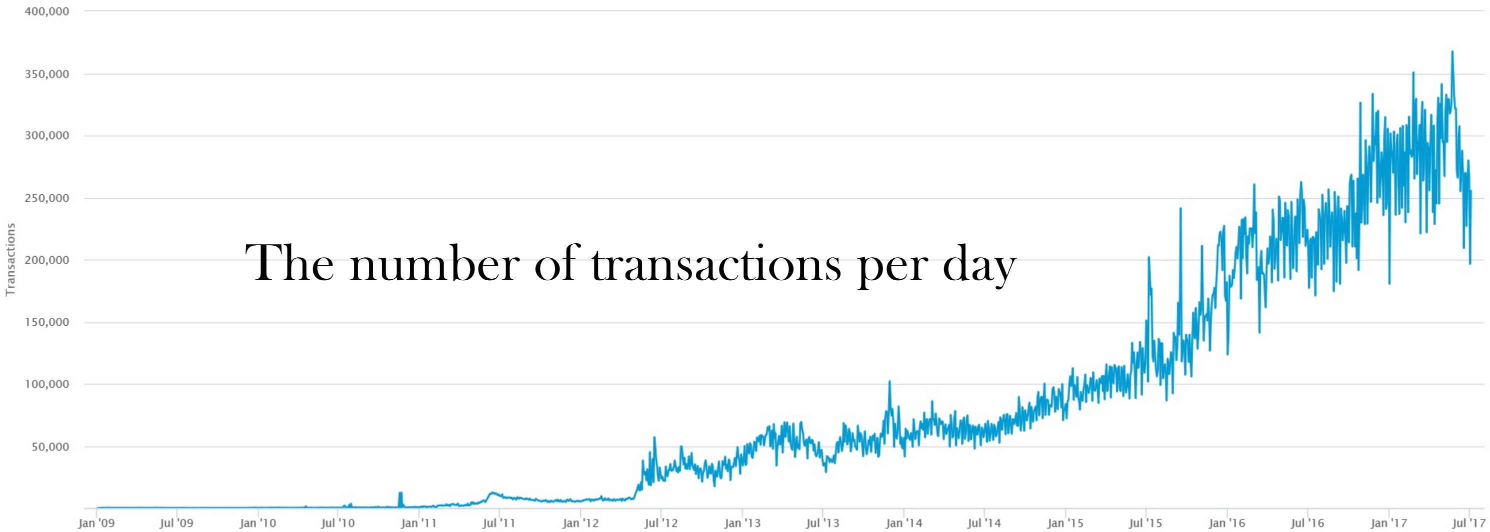
Hacking Bitcoin Mining Pool For Fun and Profit via FAW Attacks

Yongdae Kim, Yujin Kwon
Korea Advanced Institute of Science and Technology
School of Electrical Engineering
System Security Lab.

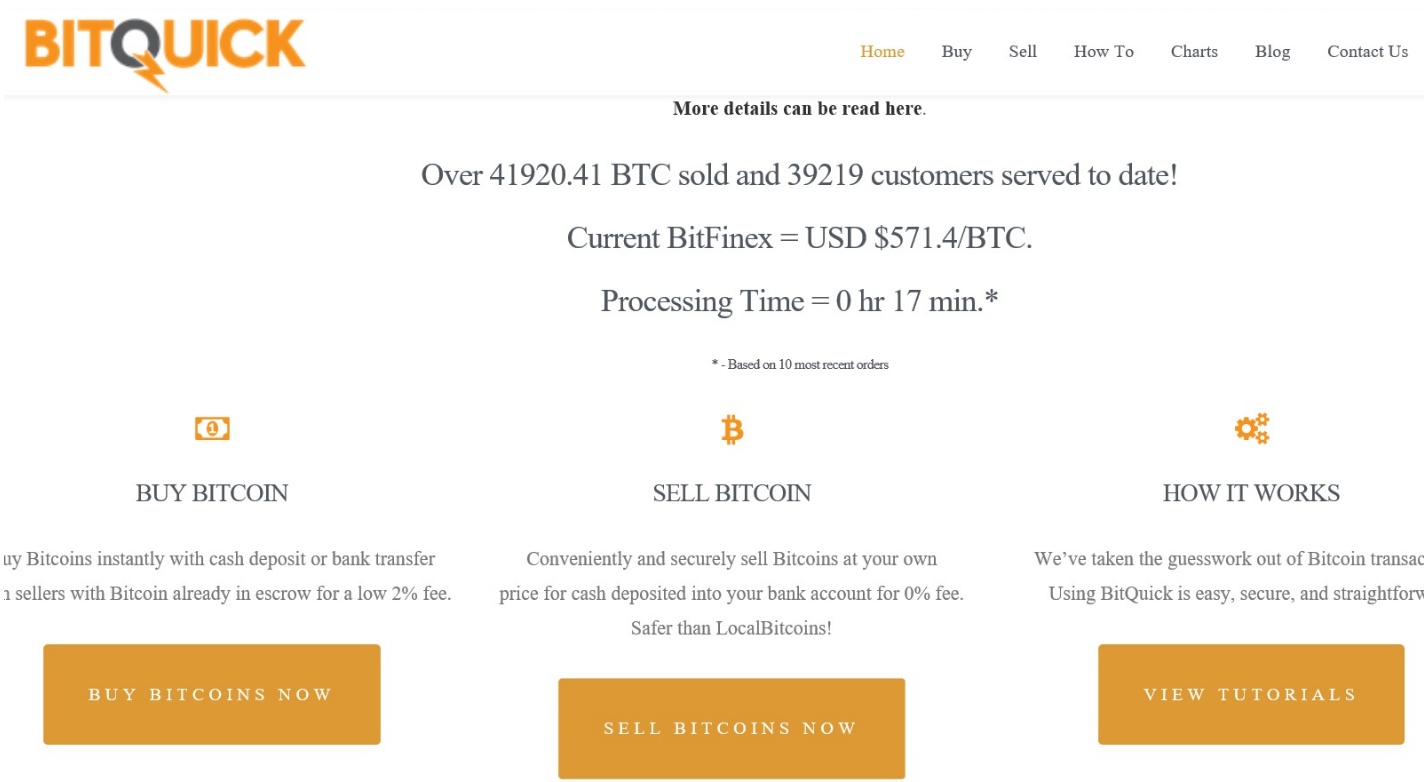
Bitcoin?

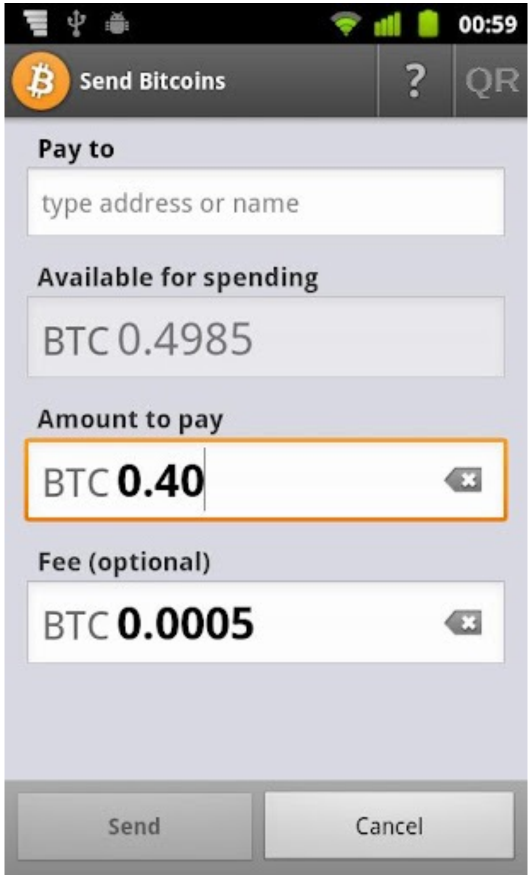
- Satoshi Nakamoto, who published the invention in 2008 and released it as open-source software in 2009.
- Bitcoin is a first cryptocurrency based on a peer-to-peer network.
- Bitcoin as a form of payment for products and services has grown, and users are increasing.



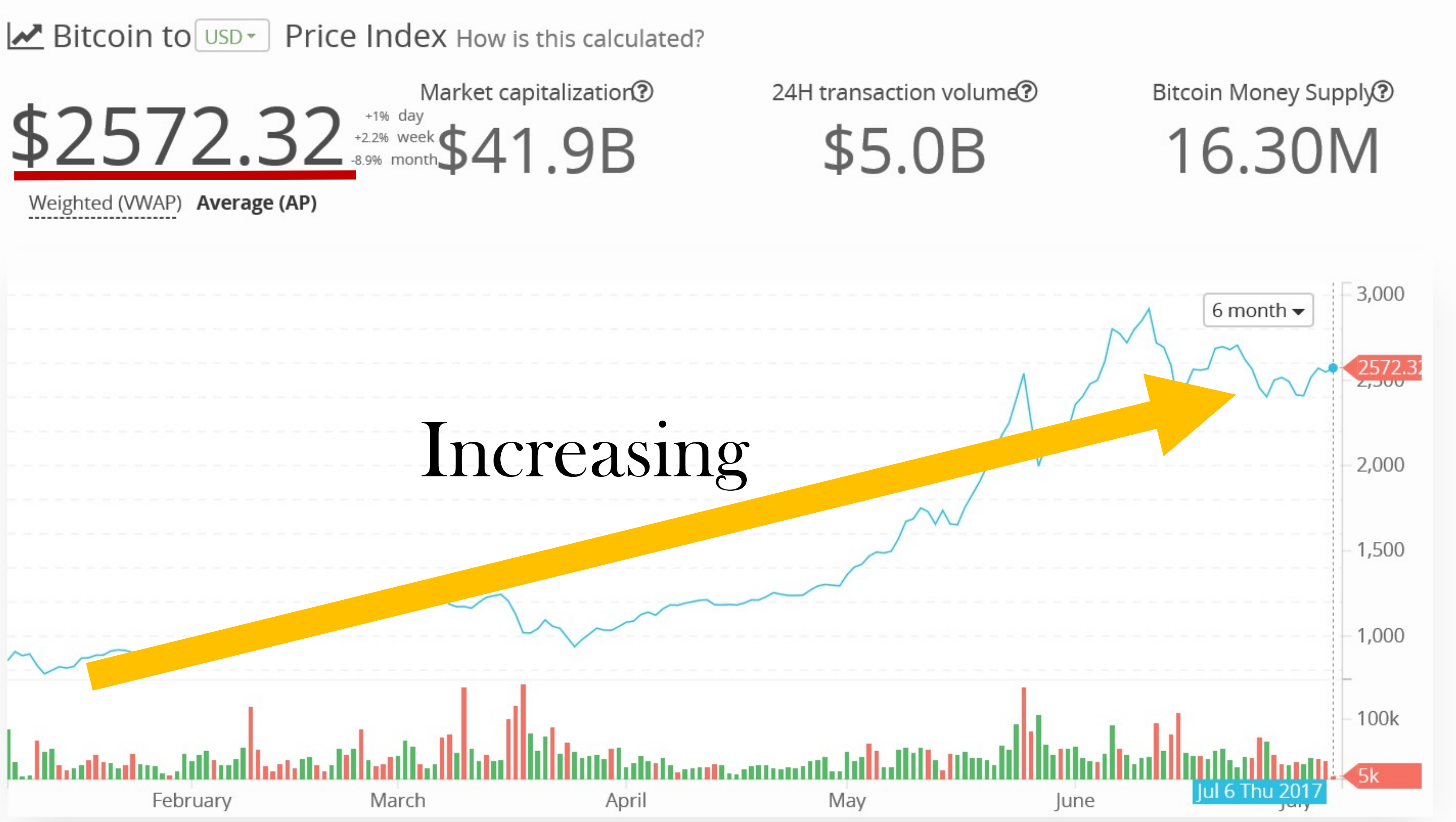
The number of transactions per day

How to Use Bitcoin

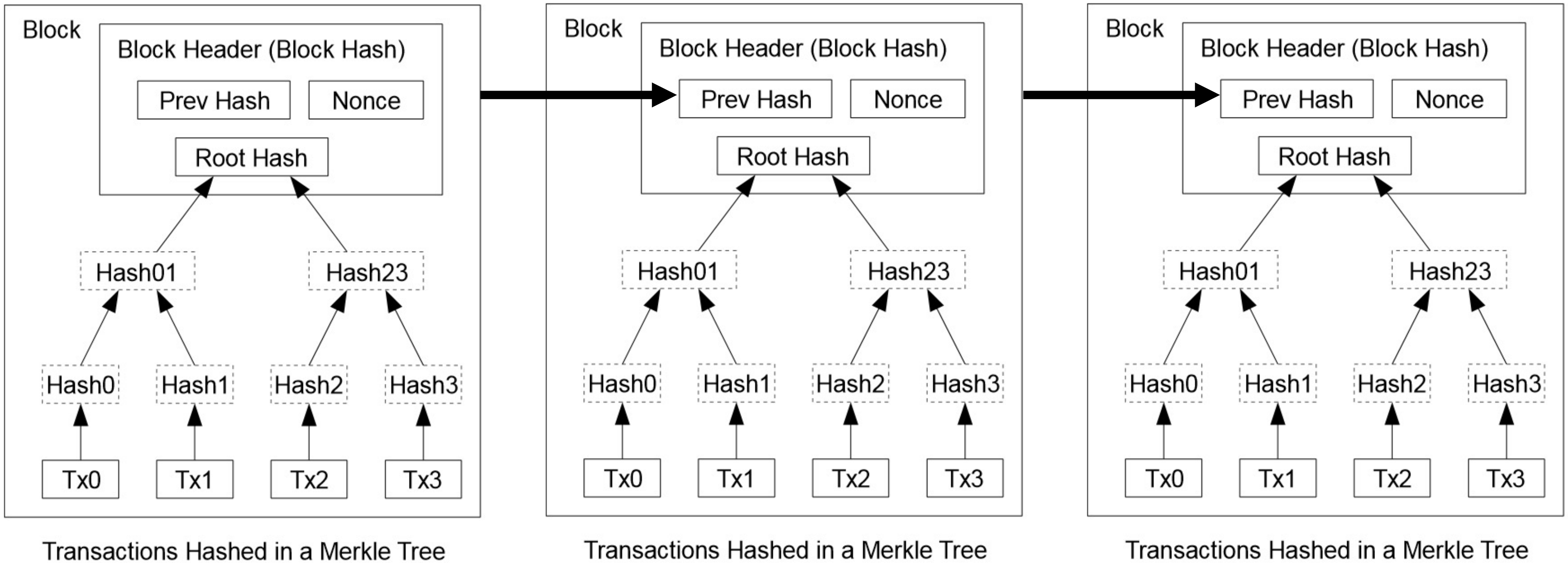




Price for 1 Bitcoin



Blockchain

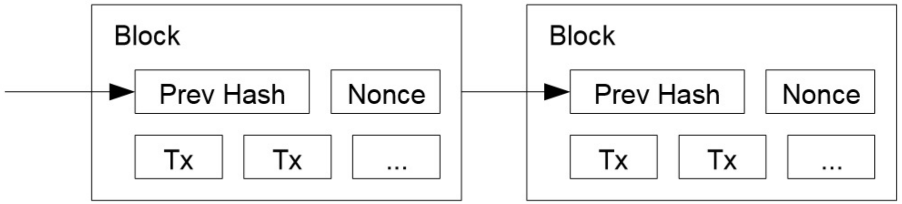


- ❖ Blocks connect as a chain.
- ❖ Each header of blocks includes the previous block's hash.

Proof-of-Work

❖ Proof-of-work scheme is based on SHA-256

❖ Proof-of-work is to find a valid Nonce by incrementing the Nonce in the block header until the block's hash value has the required prefix zero bits.



Contents

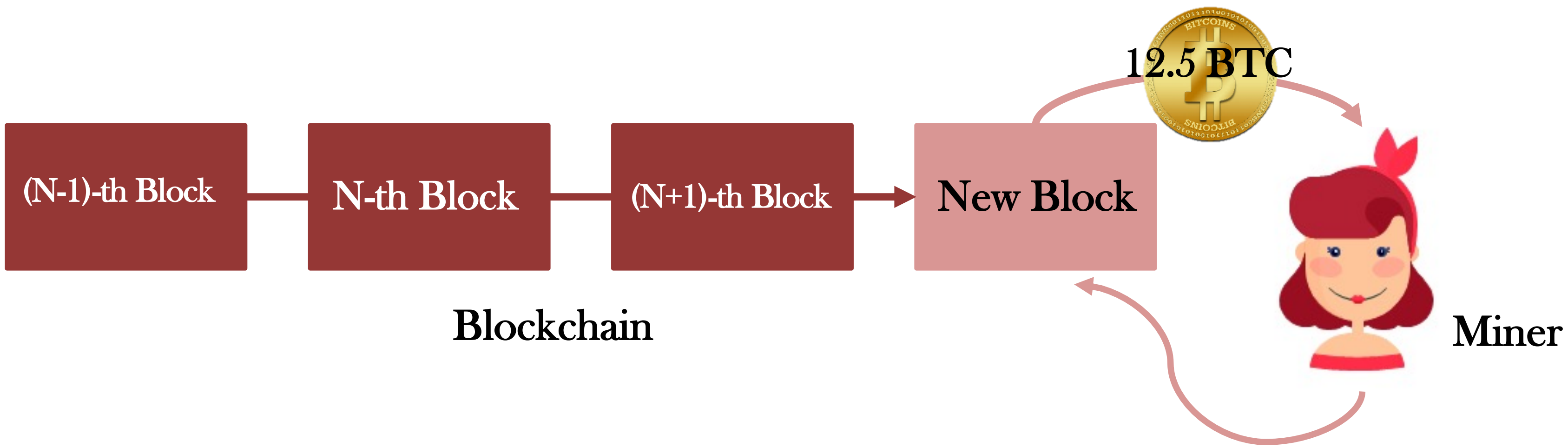
Nonce

```
"Hello, world!0" => 1312af178c253f84028d480a6adc1e25e81caa44c749ec81976192e2ec934c64
"Hello, world!1" => e9afc424b79e4f6ab42d99c81156d3a17228d6e1eef4139be78e948a9332a7d8
"Hello, world!2" => ae37343a357a8297591625e7134cbea22f5928be8ca2a32aa475cf05fd4266b7
...
"Hello, world!4248" => 6e110d98b388e77e9c6f042ac6b497cec46660deef75a55ebc7cfdf65cc0b965
"Hello, world!4249" => c004190b822f1669cac8dc37e761cb73652e7832fb814565702245cf26ebb9e6
"Hello, world!4250" => 0000c3af42fc31103f1fdc0151fa747ff87349a4714df7cc52ea464e12dcd4e9
```

Valid nonce

Reward

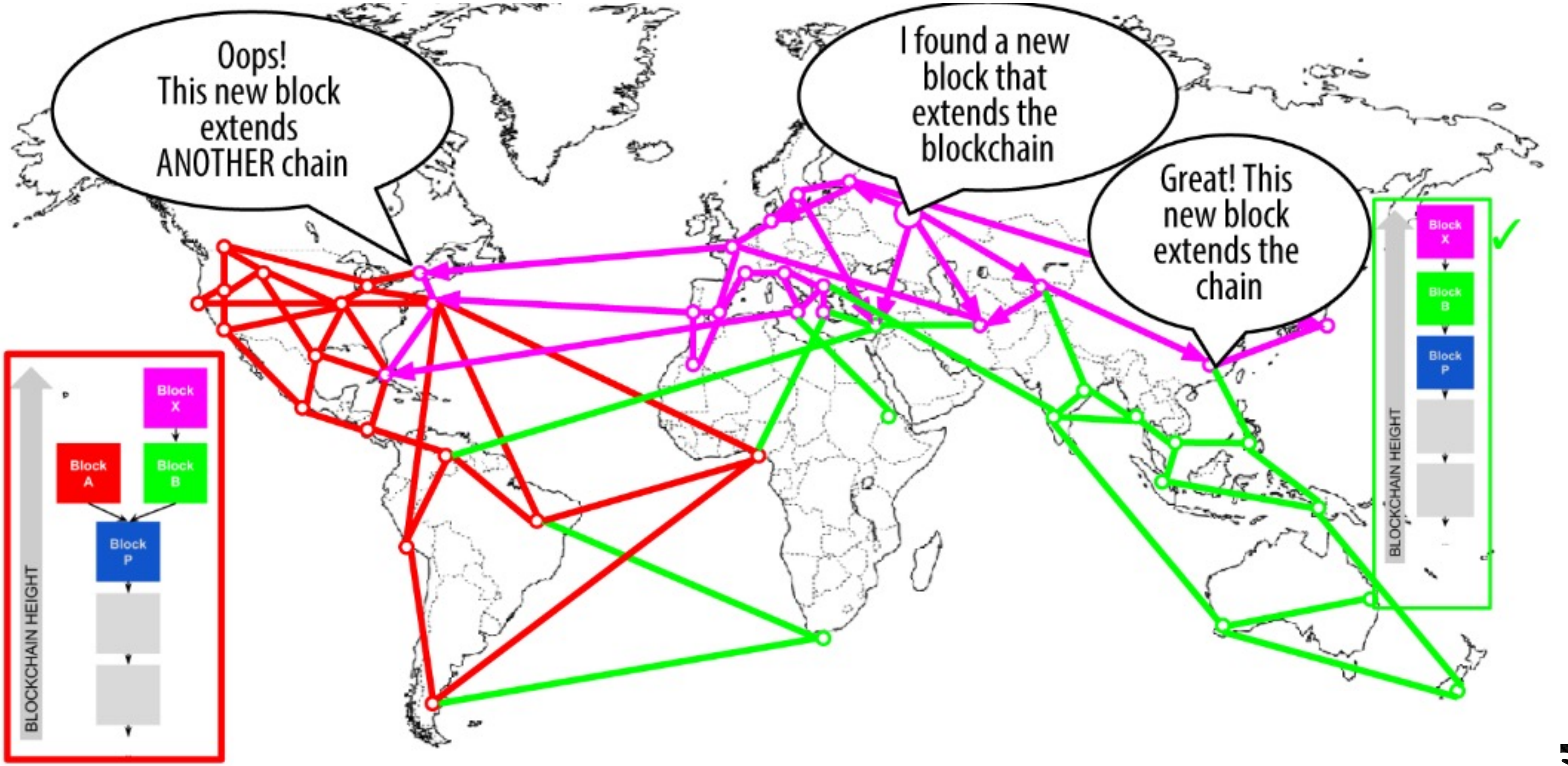
- Performing proof-of-work is called **Mining.**
- A person which do mining is called **Miner.**
- A miner can earn 12.5 BTC (≈ $ 32.5k ≈ 37M Won) as a reward when she succeeds to find a valid nonce.



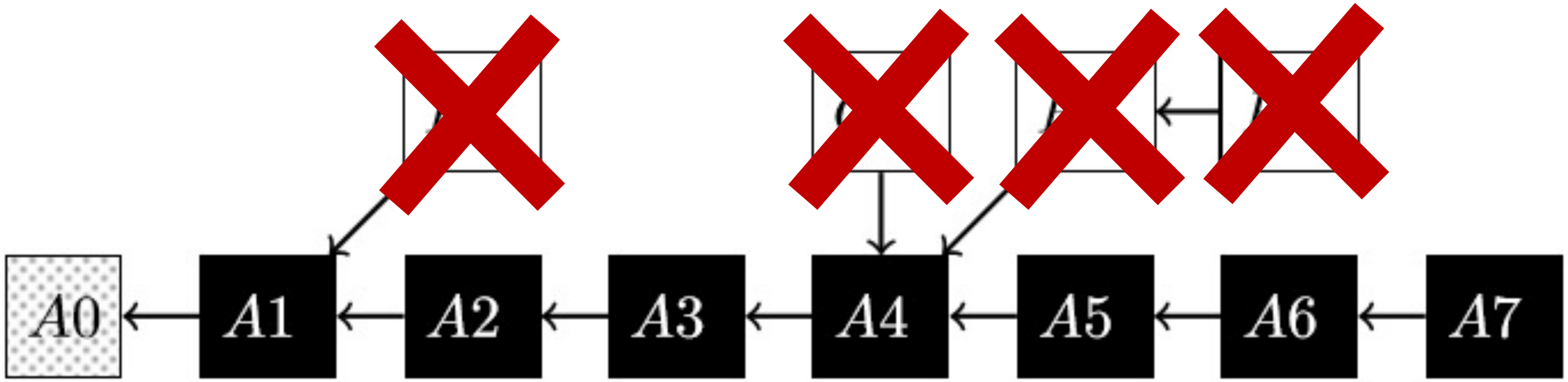
Step (Miner)

- New transactions are broadcast to all nodes.
- Each node collects new transactions into a block.
- Each node works on finding a difficult proof-of-work for its block.
- When a node finds a proof-of-work, it broadcasts the block to all nodes.
- Nodes express their acceptance of the block by working on creating the next chain, using the hash of the accepted block as the previous hash.

Forks


Oops!
This new block
extends
ANOTHER chain
I found a new
block that
extends the
blockchain
Great! This
new block
extends the
chain
Block X
Block A
Block B
Block P
BLOCKCHAIN HEIGHT
Block X
Block B
Block P
BLOCKCHAIN HEIGHT

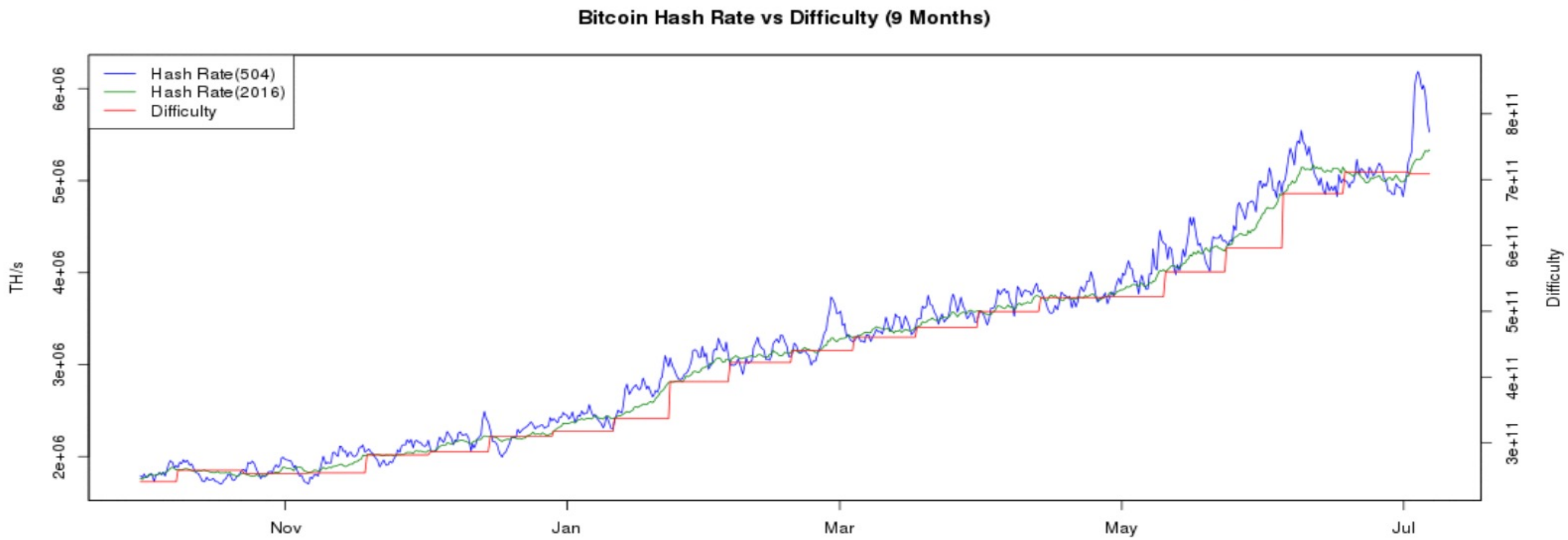
Forks



❖ Only one head is accepted as a valid one among heads.

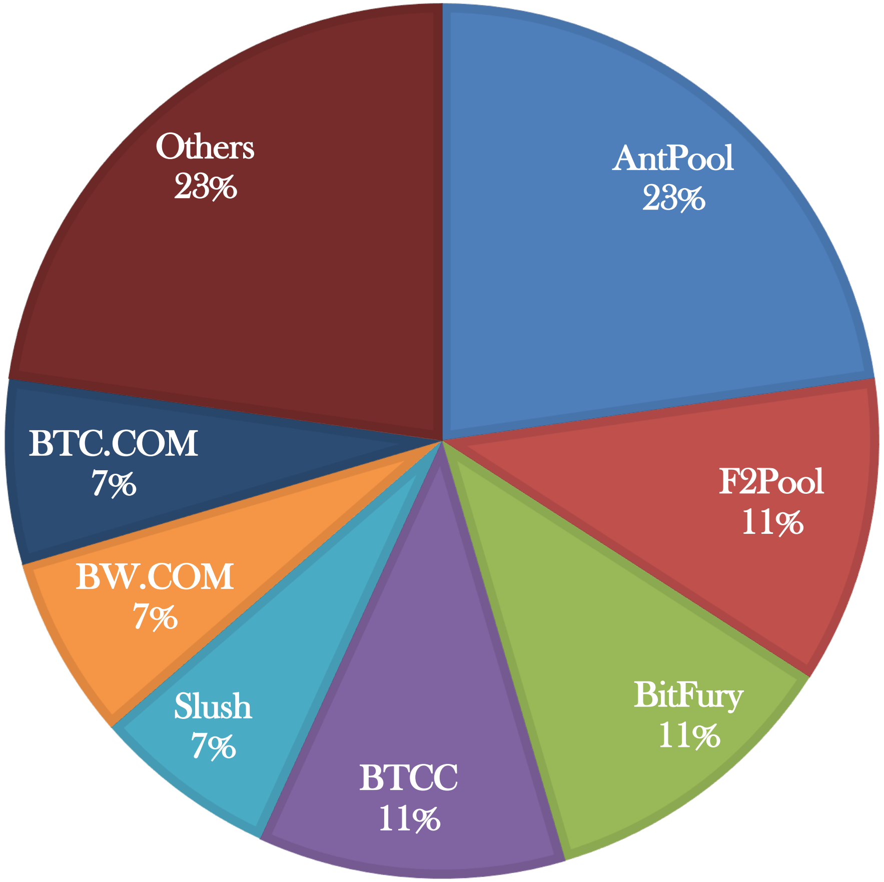
❖ An attacker can generate forks intentionally by holding his found block for a while.

Mining Difficulty



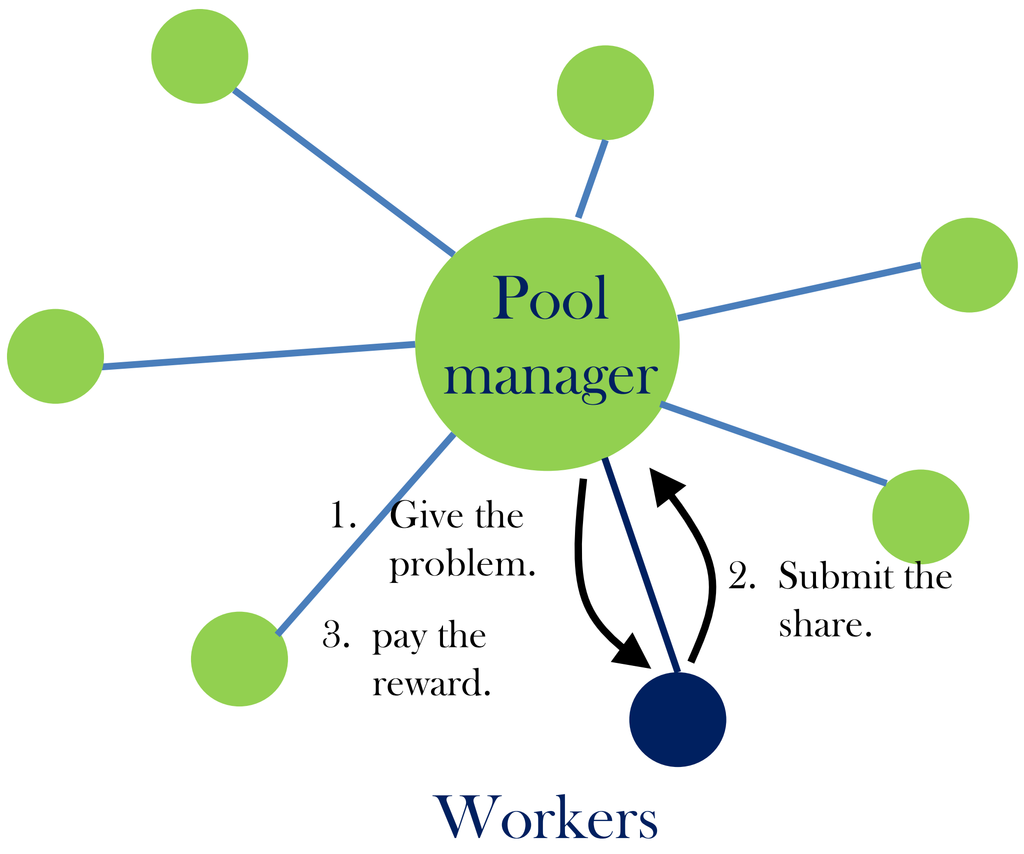
- ❖ Bitcoin adjusts automatically the mining difficulty to be an average one round period 10mins.
- ❖ The difficulty increases continuously as computing power increases.

Mining Pool



- Many miners started to do mining together.
- Most mining pools consist of a manager and miners.
- Currently, most computational power is possessed in mining pools.

Stratum

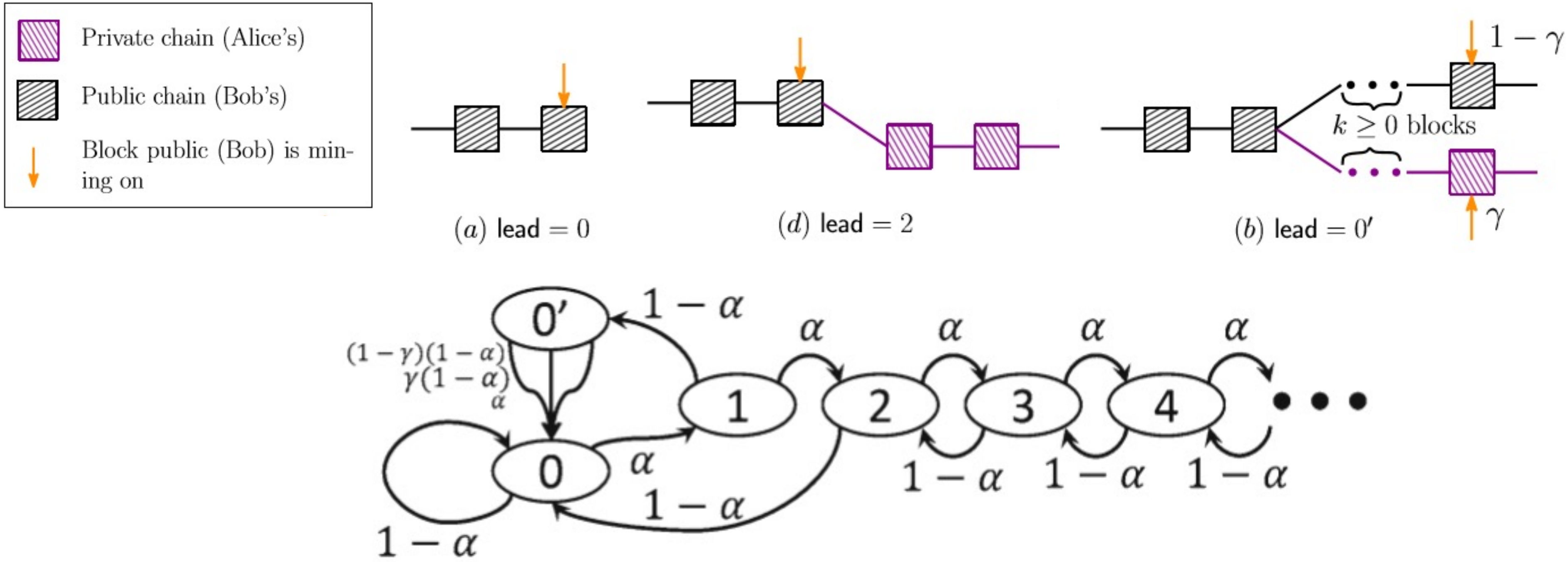


- A miner in a pool solves the easier problem than actual proofs-of-work.
- A miner submits the solution called a share to a manager.
- The manager pay the profit to a miner **in proportion to an amount of shares (easier problems solved).**

Attacks in Bitcoin System

- Double spending
- Anonymity
- Peer-to-Peer Network
- Mining
  - Selfish mining: FC 2014
    - Generate intentional forks
  - Block withholding (BWH) attacks: S&P 2015
    - Exploit pools' protocol
  - Fork after withholding (FAW) attacks
    - Generate intentional forks through pools

Selfish Mining



❖ Generate intentional forks adaptively.

❖ Force the honest miners into performing wasted computations on the stale public branch.

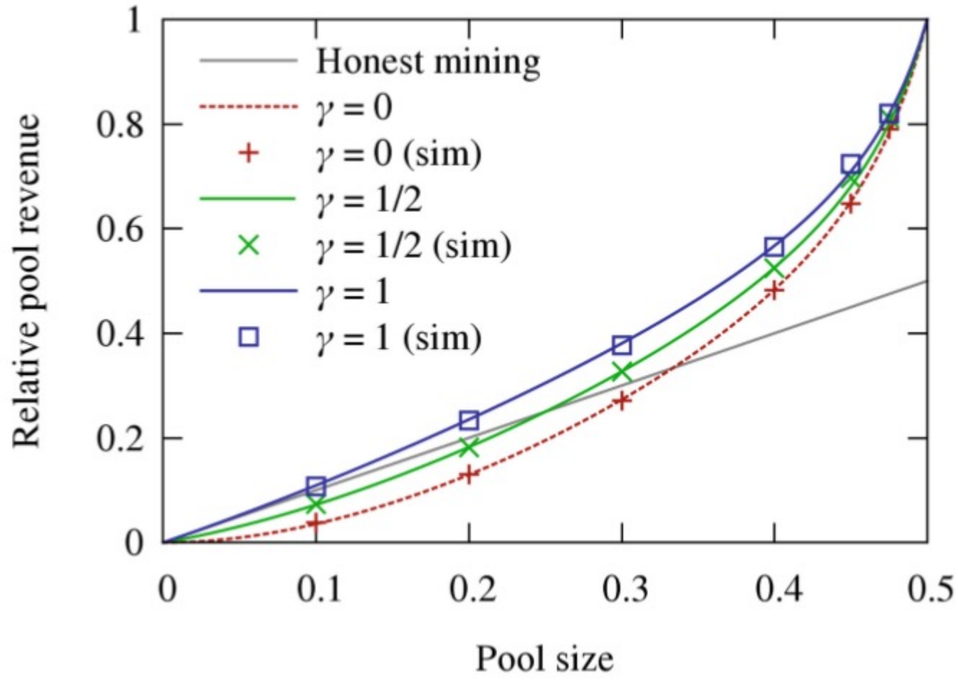
Eyal and Sirer. "Majority is not enough: Bitcoin mining is vulnerable." Financial Crypto, 2014.

Selfish Mining

❖An attacker can earn the extra reward according to her network capability.

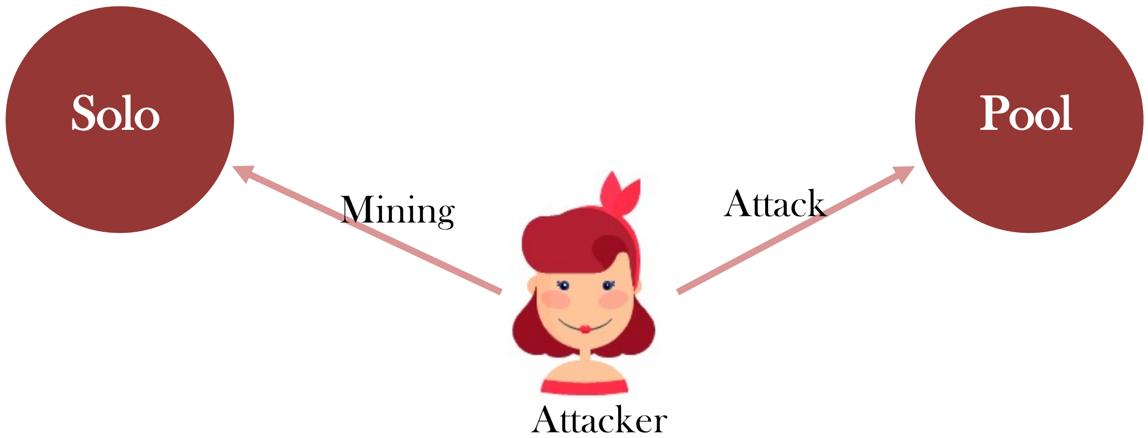
❖For example, if an attacker possesses 20% computational power, she can earn the extra reward $6M at most.

❖However, it is not practical.



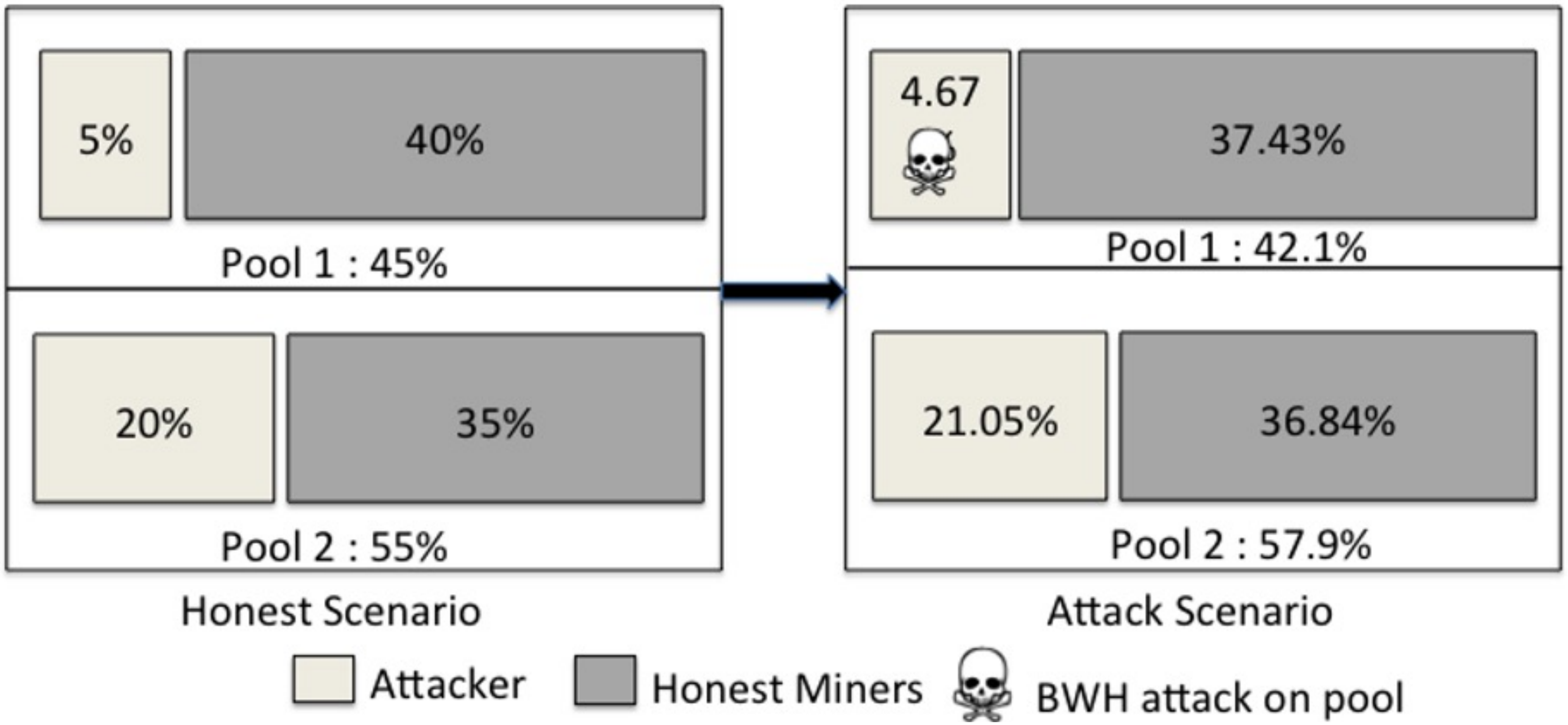
BWH Attack

- An attacker joins the target pool.
- She receives unearned wages while only pretending to contribute work in the pool.
- She submits the share which contains only partial solution but not the perfect solution.
- She should split her computational power into solo mining and malicious pool mining.



SysSec
System Security Lab

BWH Attack



5%
40%
Pool 1 : 45%
20%
35%
Pool 2 : 55%
Honest Scenario
4.67
37.43%
Pool 1 : 42.1%
21.05%
36.84%
Pool 2 : 57.9%
Attack Scenario
Attacker
Honest Miners
BWH attack on pool

FAW Att

- In the BWH a cept the target pool.
- In the FAW at vards by generating inte
- She submits or al miners propagate a blo
- For example, i she can earn the extra reward **$ 320k (≈ 369M Won)** and **$ 1053k (≈ 1215M Won)** per month via BWH and FAW attacks, respectively. (Basic reward: $ 27M ≈ 31100M Won)

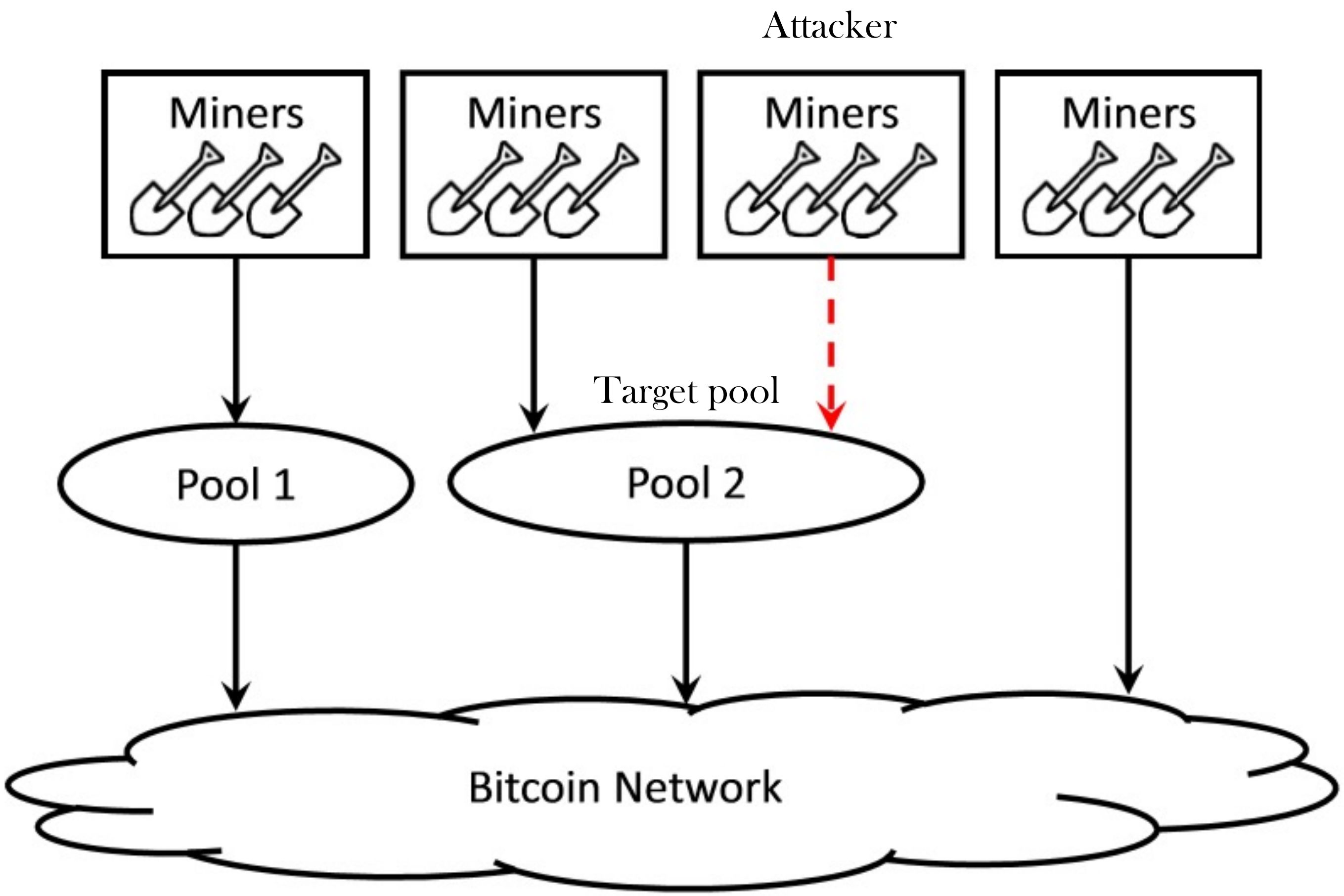
Back to the BWH Attack

The History of the BWH Attack

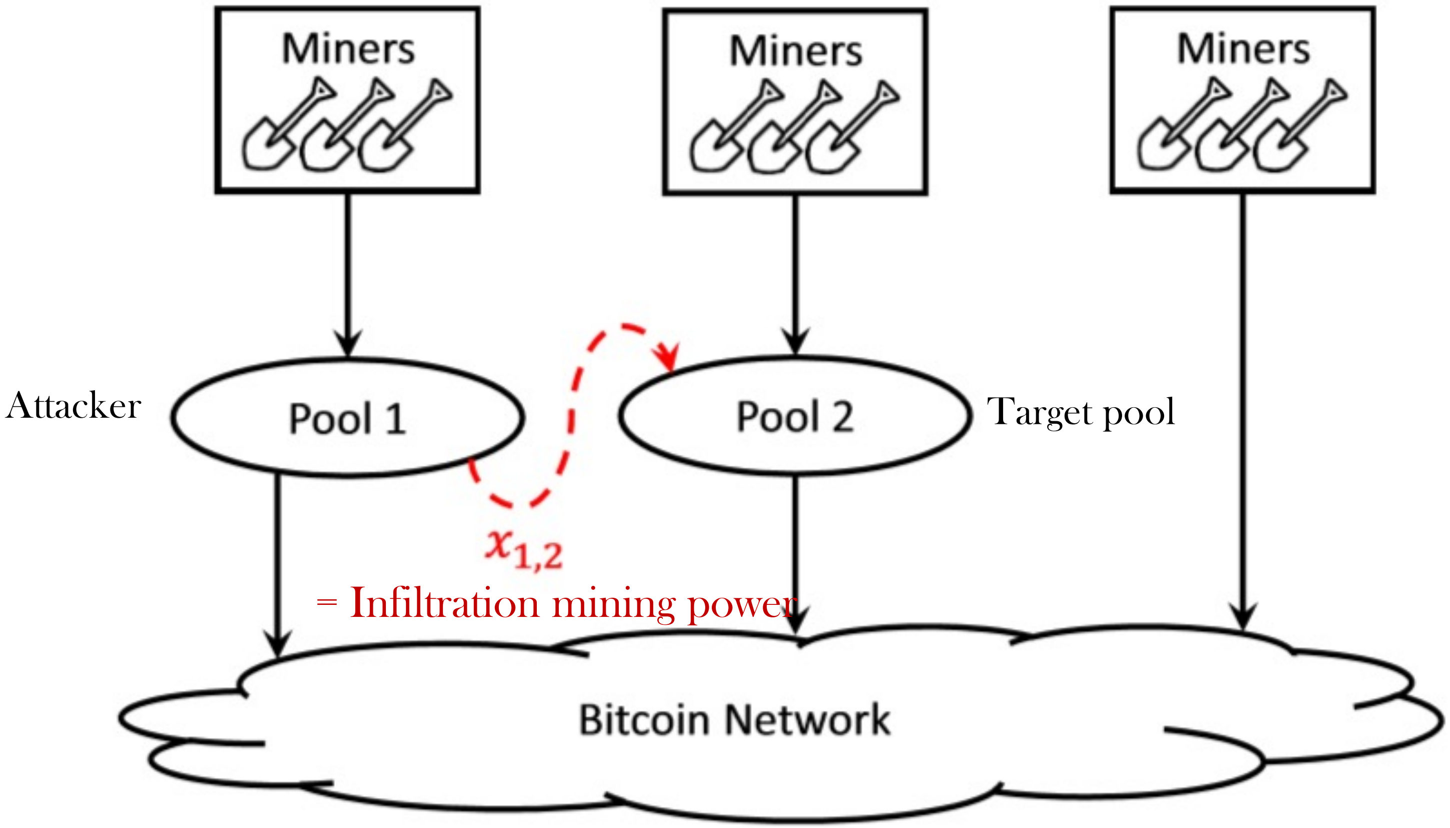
- 2011: Analysis of Bitcoin Pooled Mining Reward Systems
  - "This has no direct benefit for the attacker, only causing harm to the pool operator or participants. "
- 2014 : On Subversive Miner Strategies and Block Withholding Attack in Bitcoin Digital Currency
  - "They showed that an attacker can earn profit by this attack"
  - In june 2014, Eligius pool made a loss because of the BWH attack.
- 2015 : The miner's dilemma

  On Power Splitting Games in Distributed Computation: The Case of Bitcoin Pooled Mining
  - Attack strategy && game theory

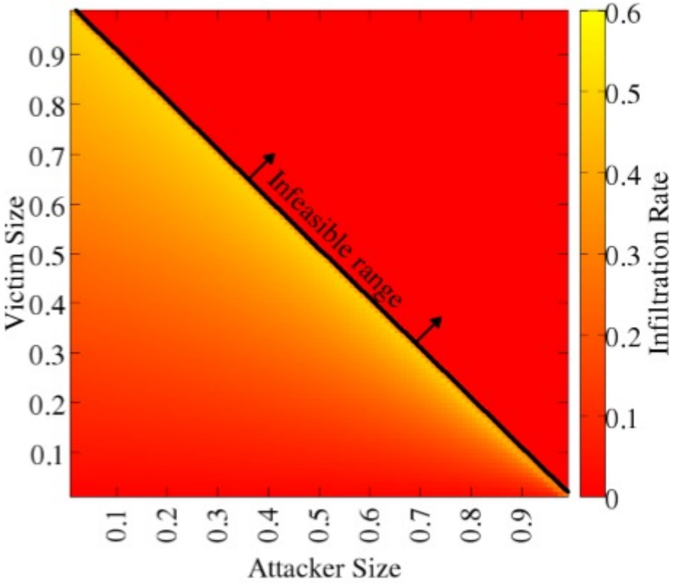
Classical BWH attack



BWH attack among pools

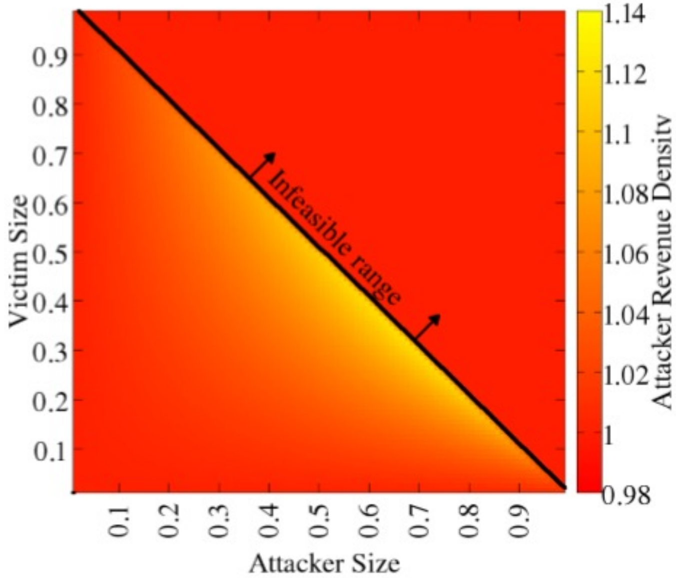


Result



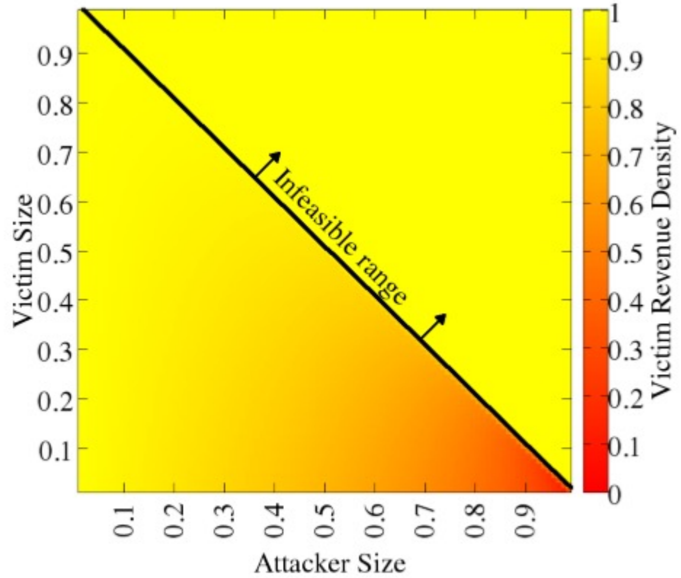
(a) $x_{1,2}$

Infiltration mining power



(b) r_1

Attacker relative reward



(c) r_2

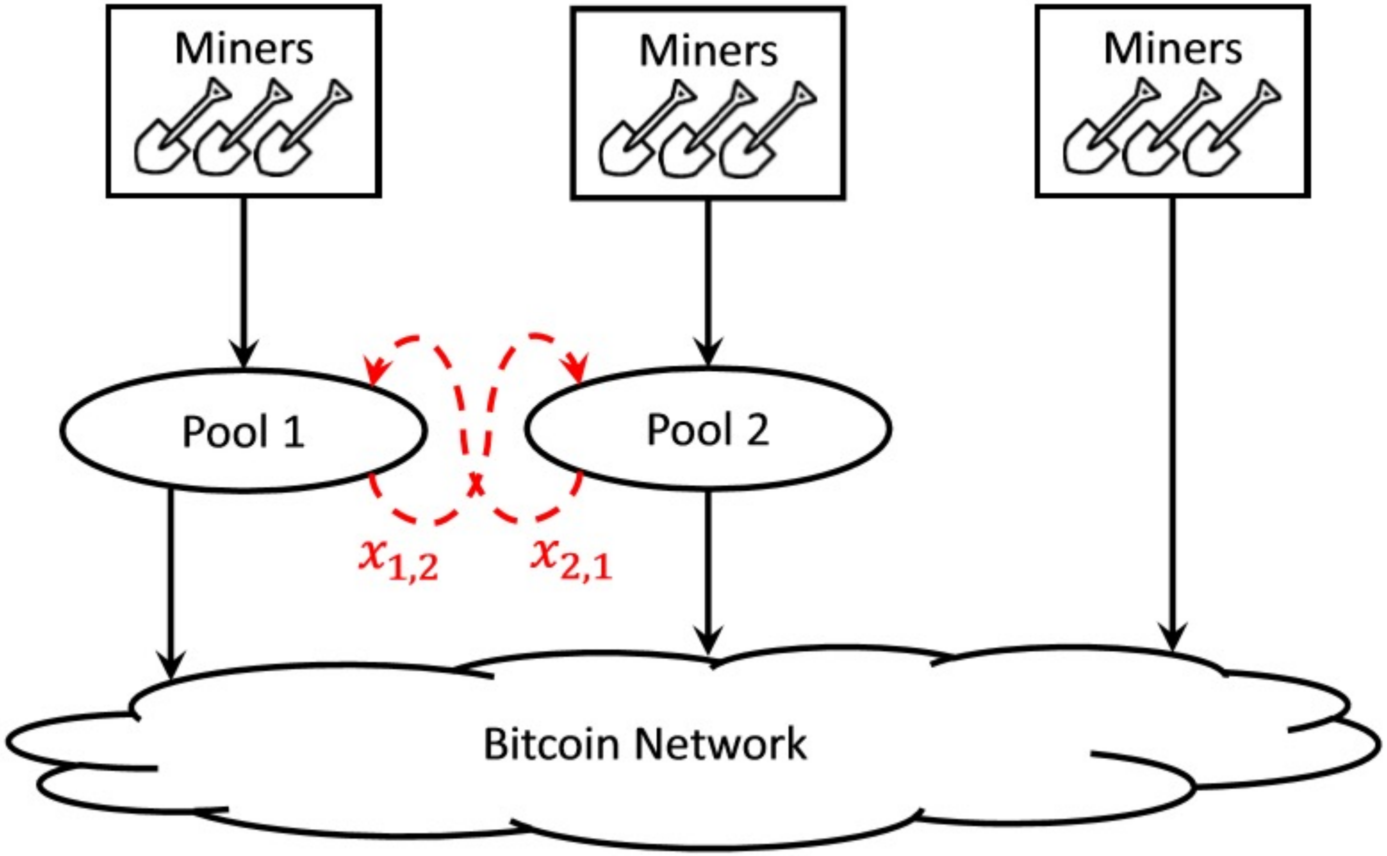
Victim relative reward

❖ The BWH attack is always profitable.

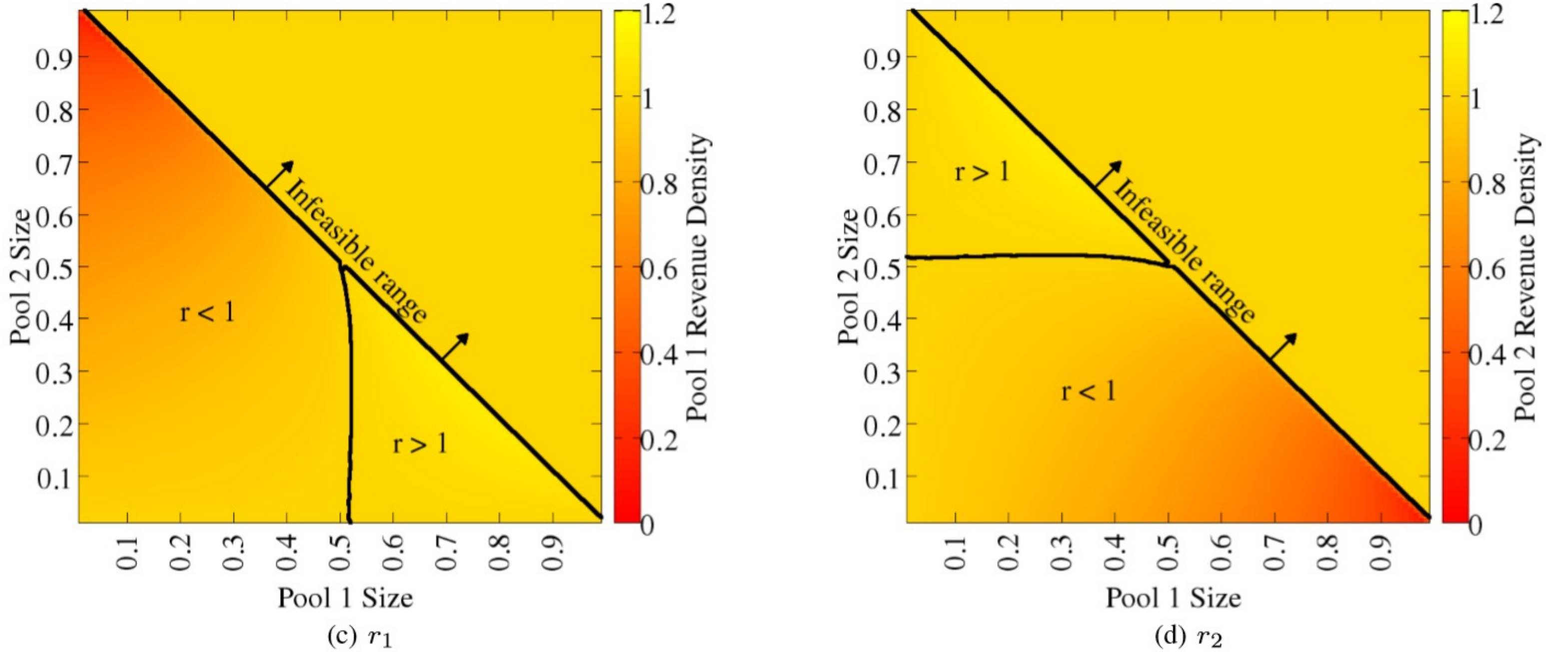
Between Two Pools

❖ Rational two pools can launch the BWH attack each other.

❖ It leads to a BWH attack game.



Result



❖ When they executes the BWH attack each other, both of them make a loss.

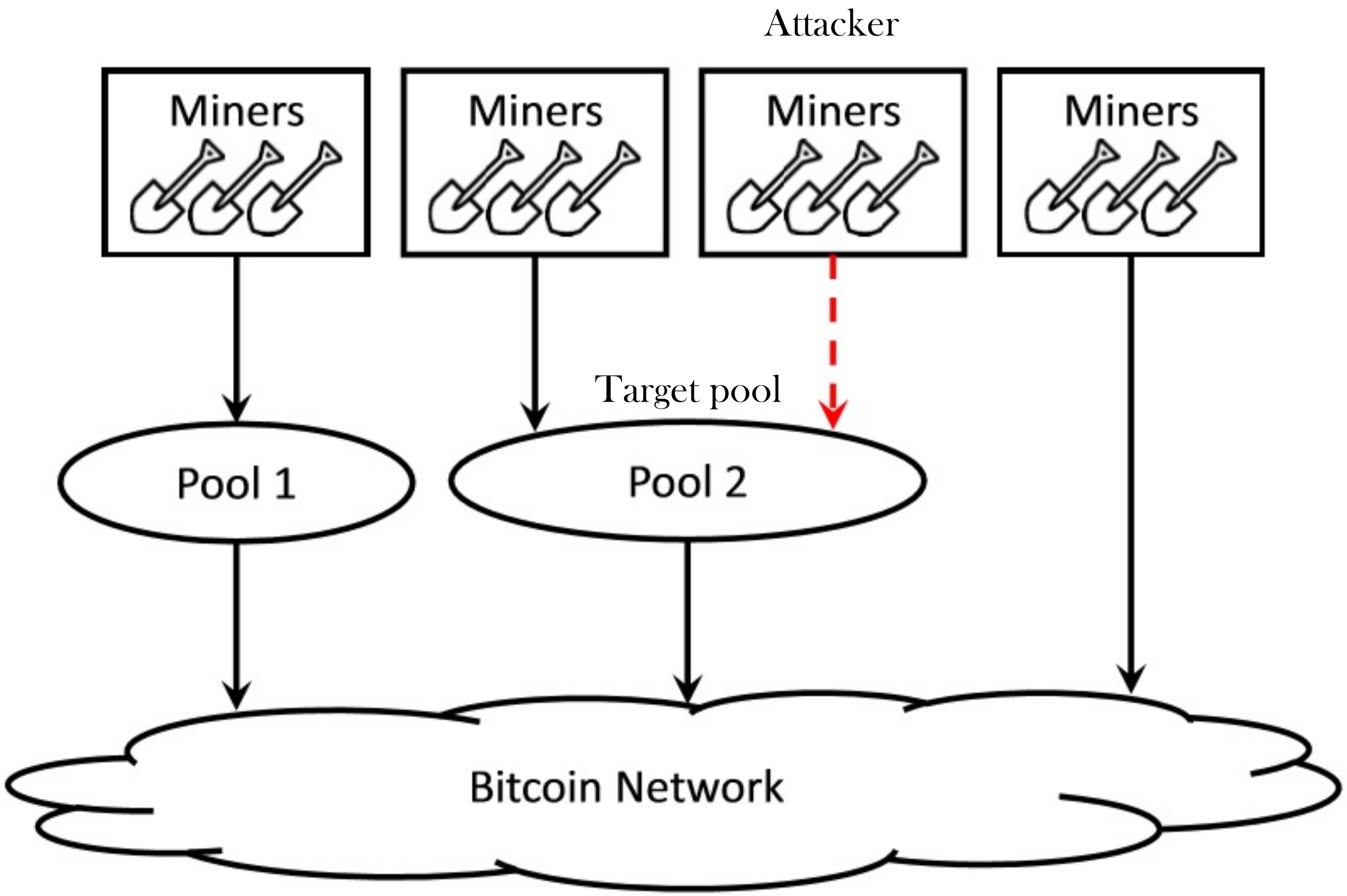
Miners' dilemma

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	$(r_1 = \tilde{r}_1 < 1, r_2 > 1)$	$(\tilde{r}_1 < r_1 < 1, \tilde{r}_2 < r_2 < 1)$

- The equilibrium revenue of the pool is inferior compared to the no-pool attacks scenario.
- This is equivalent to the prisoner's dilemma.
- The fact that the BWH attack is not common may be explained by modeling the attack decisions as an iterative prisoner's dilemma.

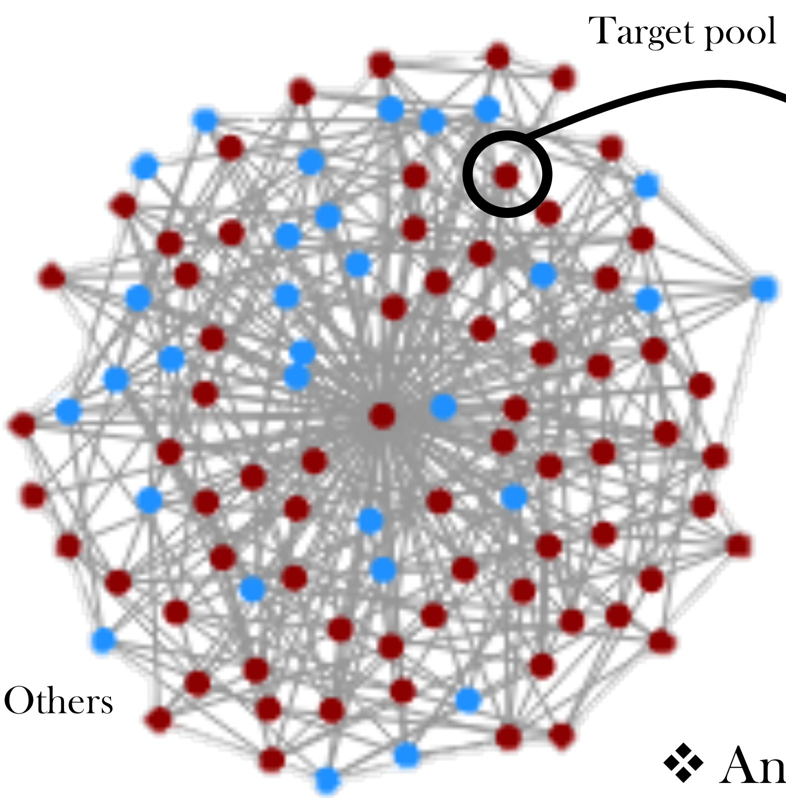
Do exist an attack which breaks the dilemma? FAW Attack

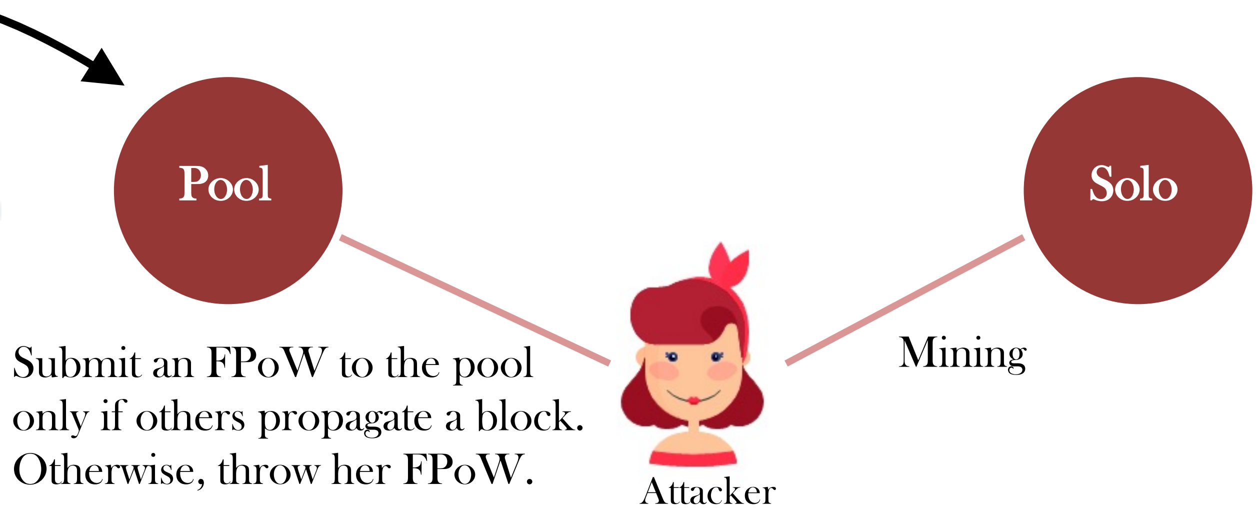
FAW Attack



Attacker
Miners
Miners
Miners
Miners
Target pool
Pool 1
Pool 2
Bitcoin Network

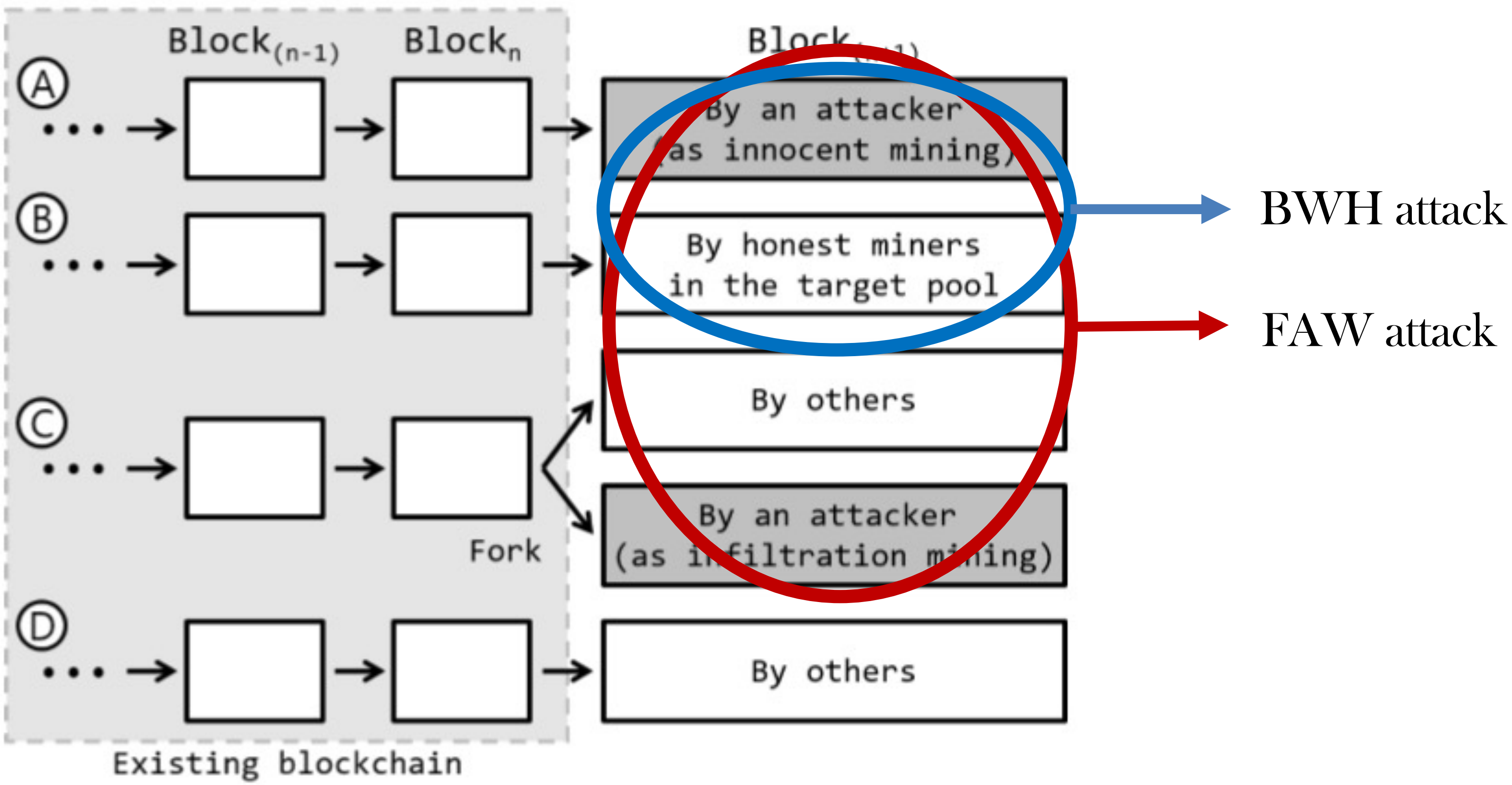
FAW Attack





❖ An attacker generates forks intentionally through a pool!

FAW Attack Against One Pool



FAW Attack Against One Pool

- Notation
  - α: Computational power of the attacker
  - β: Total computational power of a victim pool
  - γ: The infiltration mining power divided by α
  - c: Attacker's network capability
  - R_a (R_p): An attacker's (The victim's) reward

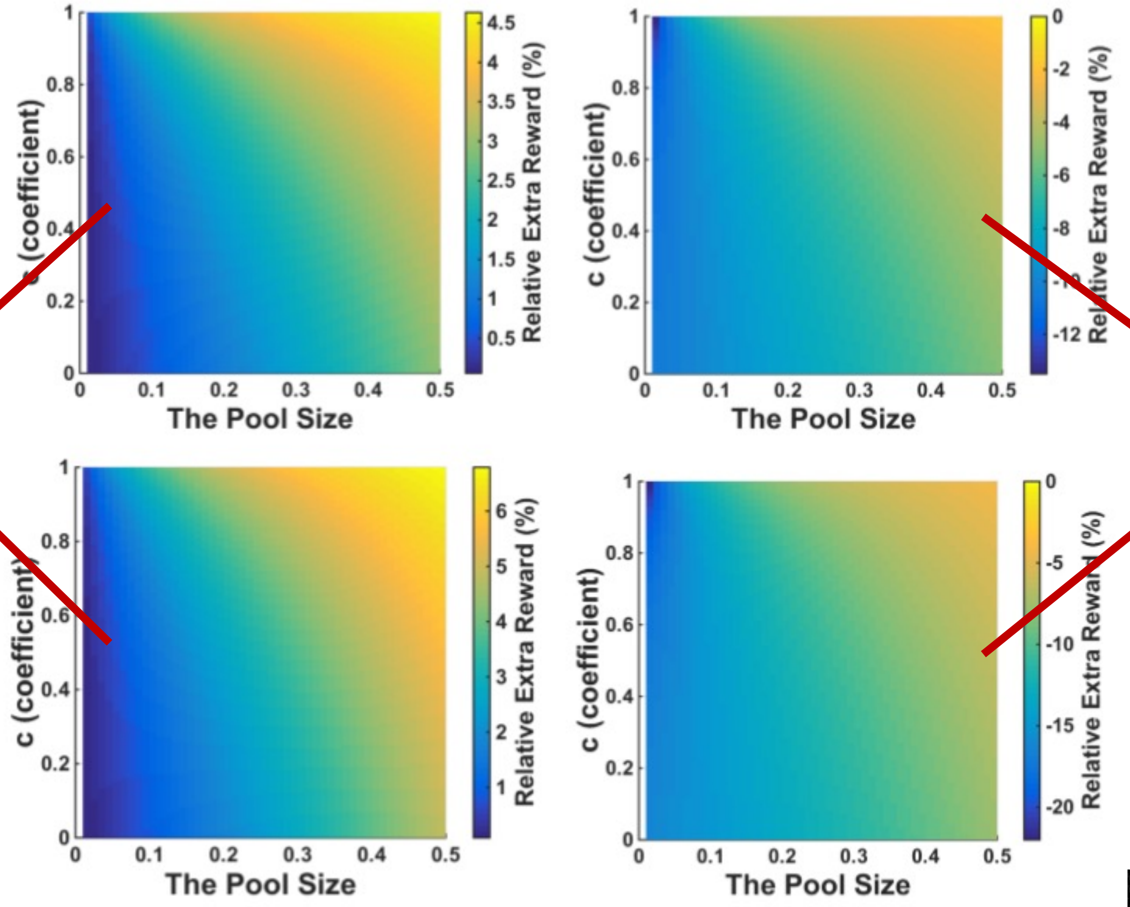
- The optimal infiltration mining power is

$$\overline{\gamma} = \frac{(1-\alpha)(1-c)\beta + \beta^2 c - \beta\sqrt{(1-\alpha-\beta)^2c^2 + ((1-\alpha-\beta)(\alpha\beta+\alpha-2))c - \alpha(1+\beta)+1}}{\alpha(1-\alpha-\beta)(c(1-\beta)-1)}$$

- The FAW attack is always profitable.

Result

Attacker　　　　Victim



Always positive

Always negative

❖ An attacker with 0.2 power

❖ An attacker with 0.3 power

Result

Increasing

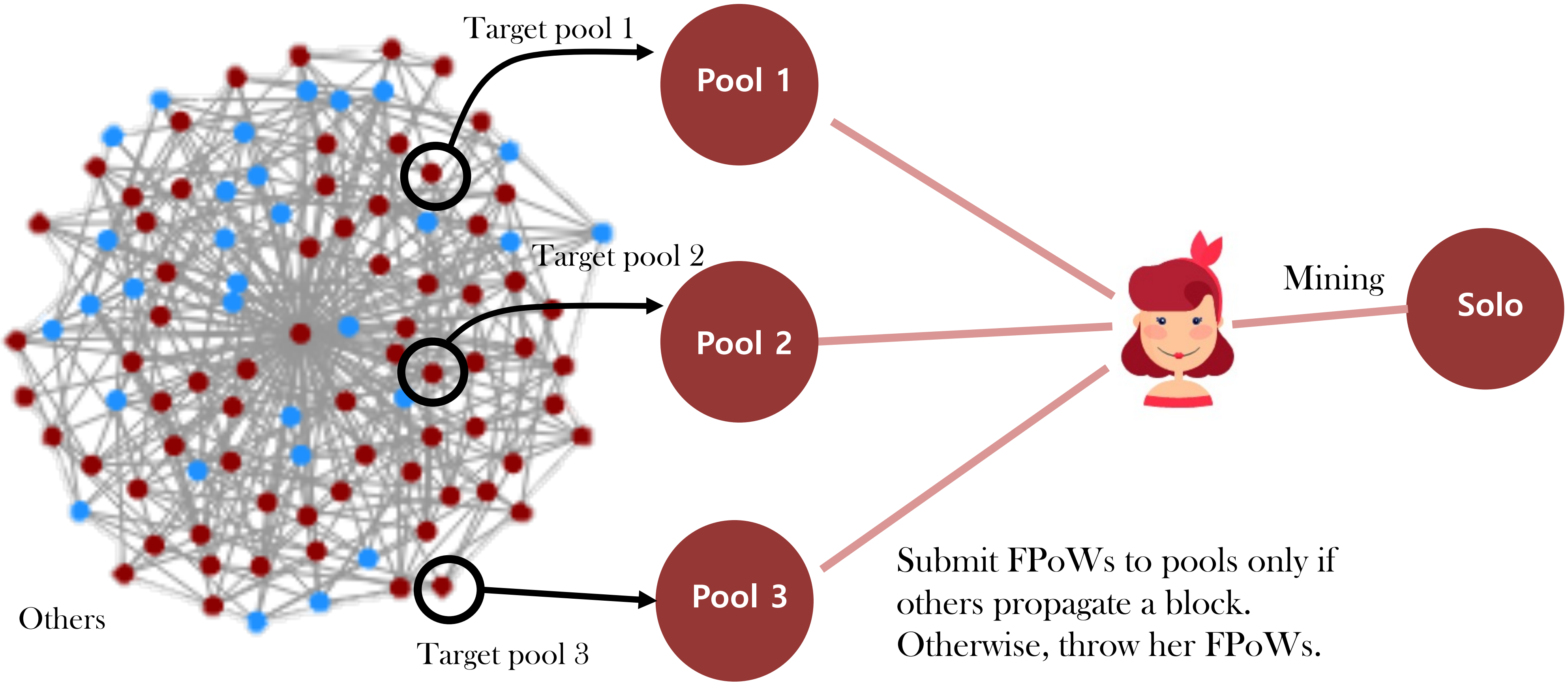
c \ α	**0.1**	**0.2**	**0.3**	**0.4**
0	0.53 (0.53)	1.14 (1.14)	1.85 (1.85)	2.70 (2.70)
0.25	0.65 (0.67)	1.38 (1.38)	2.20 (2.20)	3.1 (3.13)
0.5	0.85 (0.85)	1.74 (1.74)	2.70 (2.70)	3.75 (3.75)
0.75	1.21 (1.22)	2.37 (2.37)	3.52 (3.52)	4.69 (4.70)
1	2.12 (2.12)	3.75 (3.75)	5.13 (5.13)	6.37 (6.36)

The case is equivalent to the case of the BWH attack

Increasing

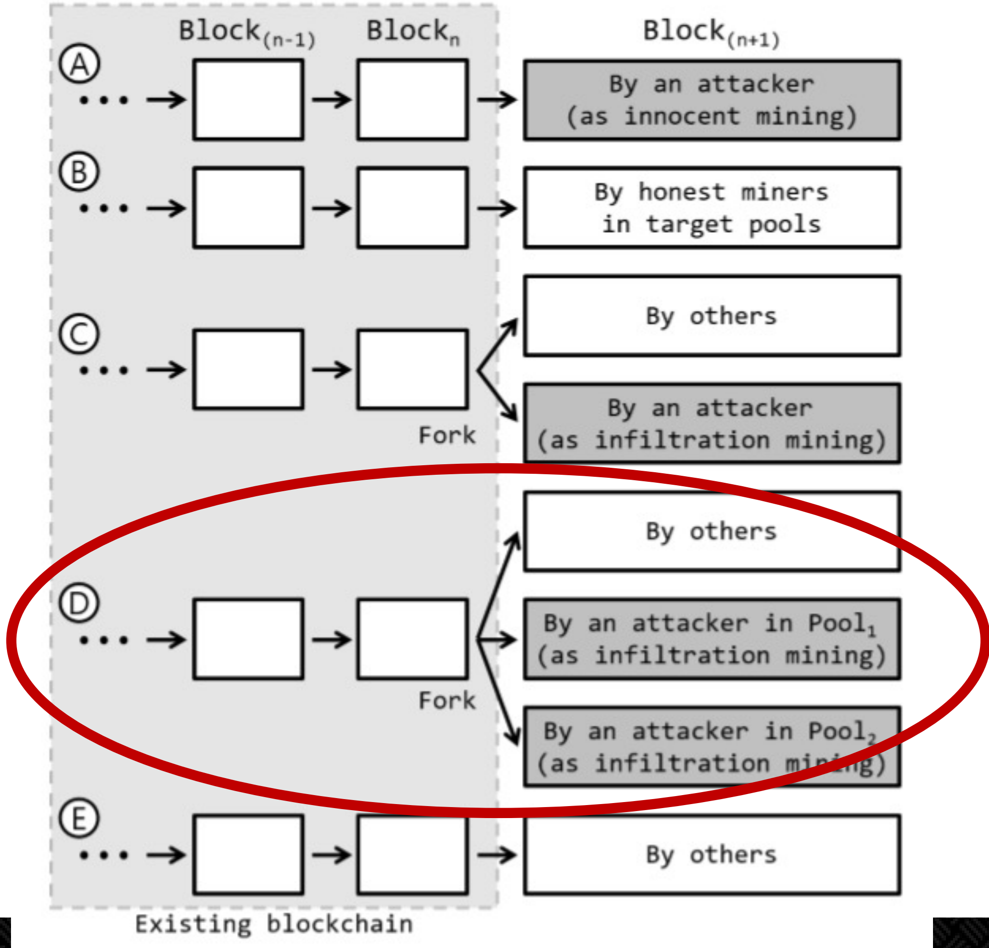
- We simulated an FAW attack against one pool which possesses a computational power of 0.2, using a Monte Carlo method.

FAW Attack Against Multiple Pools



Submit FPoWs to pools only if others propagate a block. Otherwise, throw her FPoWs.

FAW Attack Against Two Pools



FAW Attack Against Multiple Pools

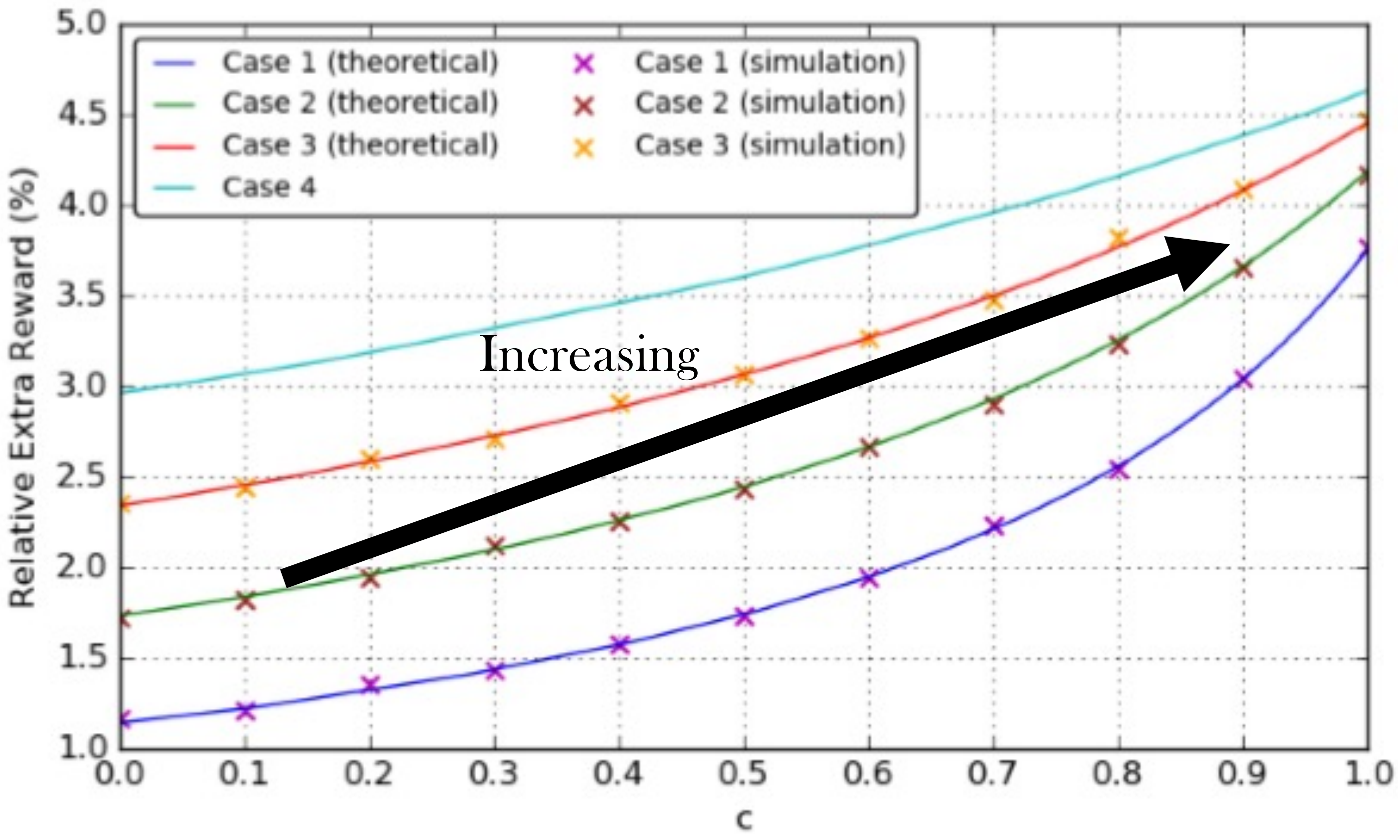
- An attacker's reward R_a is

$$R_a = \frac{(1-\gamma_1-\gamma_2)\alpha}{1-(\gamma_1+\gamma_2)\alpha} + \sum_{i=1,2}\{(\frac{\beta_i}{1-(\gamma_1+\gamma_2)\alpha} + c_i^{(1)}\gamma_i\alpha\frac{1-\alpha-\beta_1-\beta_2}{1-\gamma_i\alpha} + c_i^{(2)}\sum_j\{\gamma_j\alpha\frac{\gamma_{\neg j}\alpha}{1-\gamma_i\alpha}\}\frac{1-\alpha-\beta_1-\beta_2}{1-(\gamma_1+\gamma_2)\alpha})\cdot\frac{\gamma_i\alpha}{\beta_i+\gamma_i\alpha}\}$$

- We generalize to n target pools.

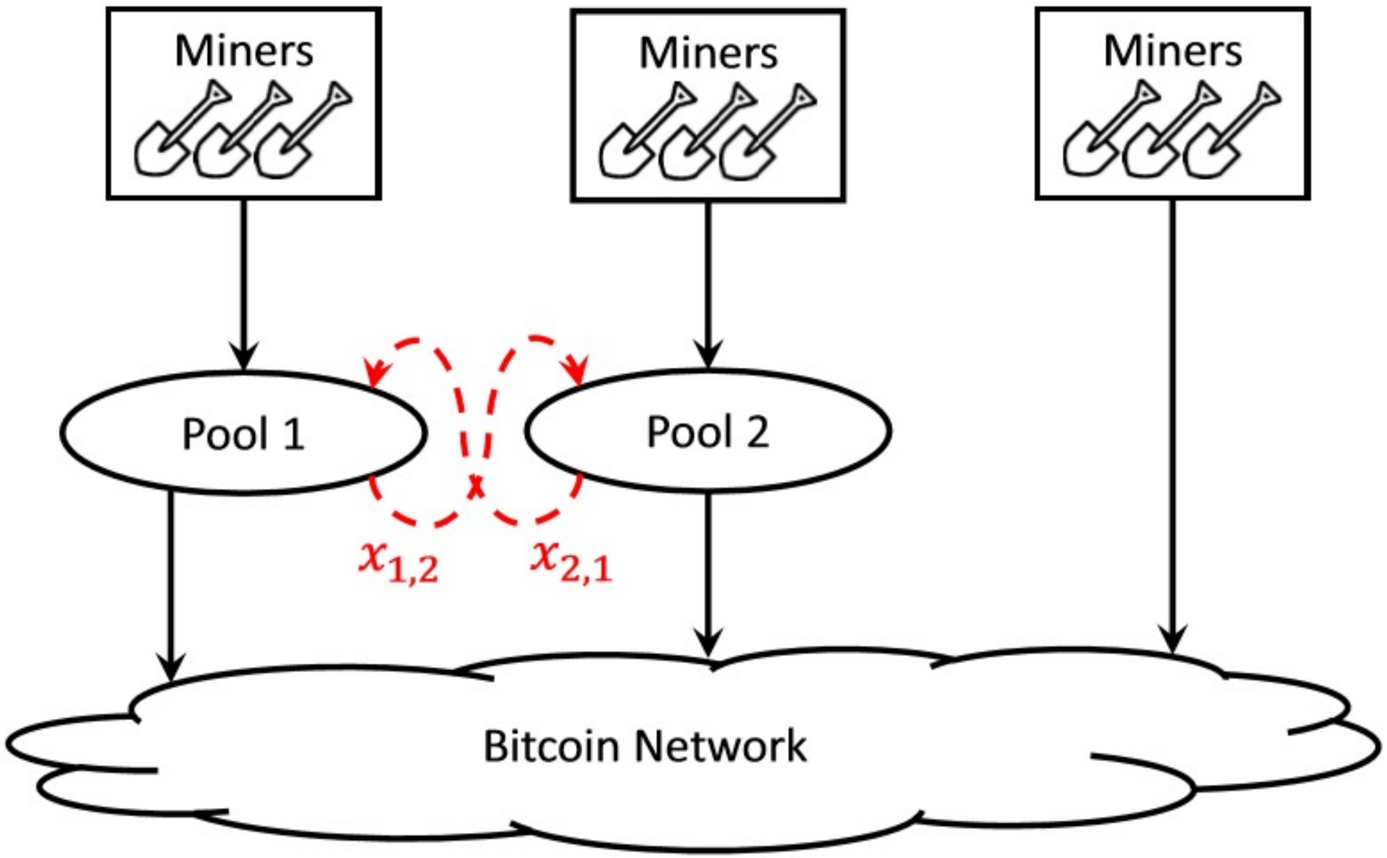
$$R_a = \frac{(1-\gamma)\alpha}{1-\gamma\alpha} + \sum_{i=1}^{n}\{(\frac{\beta_i}{1-\gamma\alpha} + \sum_{k=1}^{n}\{(1-\alpha-\beta)\sum_{\mathcal{P}_{k,i}\in\mathcal{P}}\{c_{\mathrm{Im}(\mathcal{P}_{k,i})}(i)\prod_{t=1}^{k}\frac{\gamma_{\mathcal{P}_{k,i}(t)}\alpha}{1-\sum_{d=1}^{t}\gamma_{\mathcal{P}_{k,i}(d)}\alpha}\}\})\cdot\frac{\gamma_i\alpha}{\beta_i+\gamma_i\alpha}\}$$

Result



- An attacker possesses 0.2 computational power.
- Case 1, 2, and 3 represent when two target pools' computational power (β_1, β_2) are (0.1, 0.1), (0.2, 0.1), and (0.3, 0.1), respectively.
- Case 4 considers the current power distribution. At that time, FAW attacks make her rewards greater 56% than that for BWH attacks.

FAW Attack Game

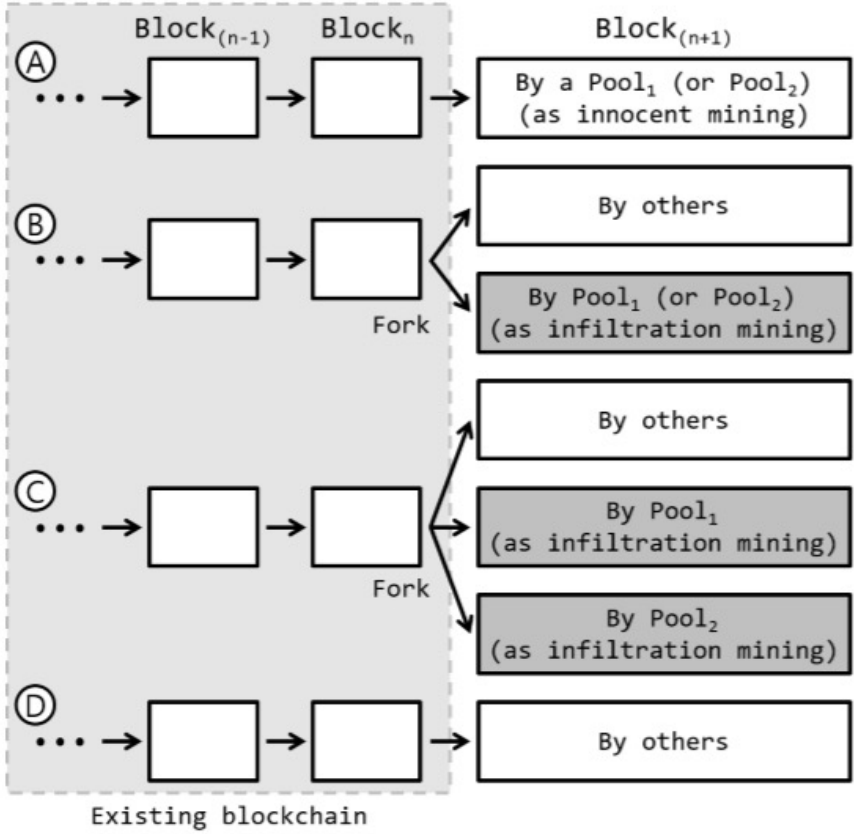


FAW Attack Game

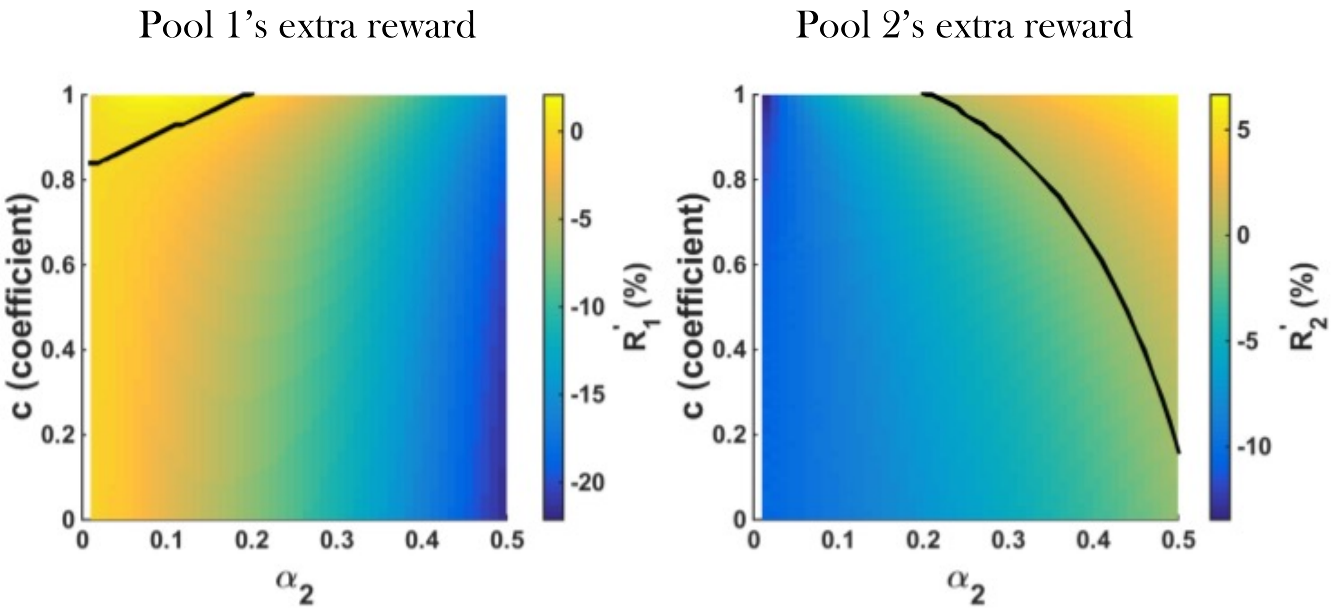
❖ Two pools attack each other. ⇒ *FAW Attack Game between two pools*

$$R_1 = \frac{\alpha_1 - f_1}{1 - f_1 - f_2} + c_2 f_2 \frac{1 - \alpha_1 - \alpha_2}{1 - f_2} + c_2^{'} f_1 f_2 \left(\frac{1}{1 - f_1} + \frac{1}{1 - f_2}\right) \frac{1 - \alpha_1 - \alpha_2}{1 - f_1 - f_2} + R_2 \frac{f_1}{\alpha_2 + f_1}$$

$$R_2 = \frac{\alpha_2 - f_2}{1 - f_1 - f_2} + c_1 f_1 \frac{1 - \alpha_1 - \alpha_2}{1 - f_1} + c_1^{'} f_1 f_2 \left(\frac{1}{1 - f_1} + \frac{1}{1 - f_2}\right) \frac{1 - \alpha_1 - \alpha_2}{1 - f_1 - f_2} + R_1 \frac{f_2}{\alpha_1 + f_2}$$

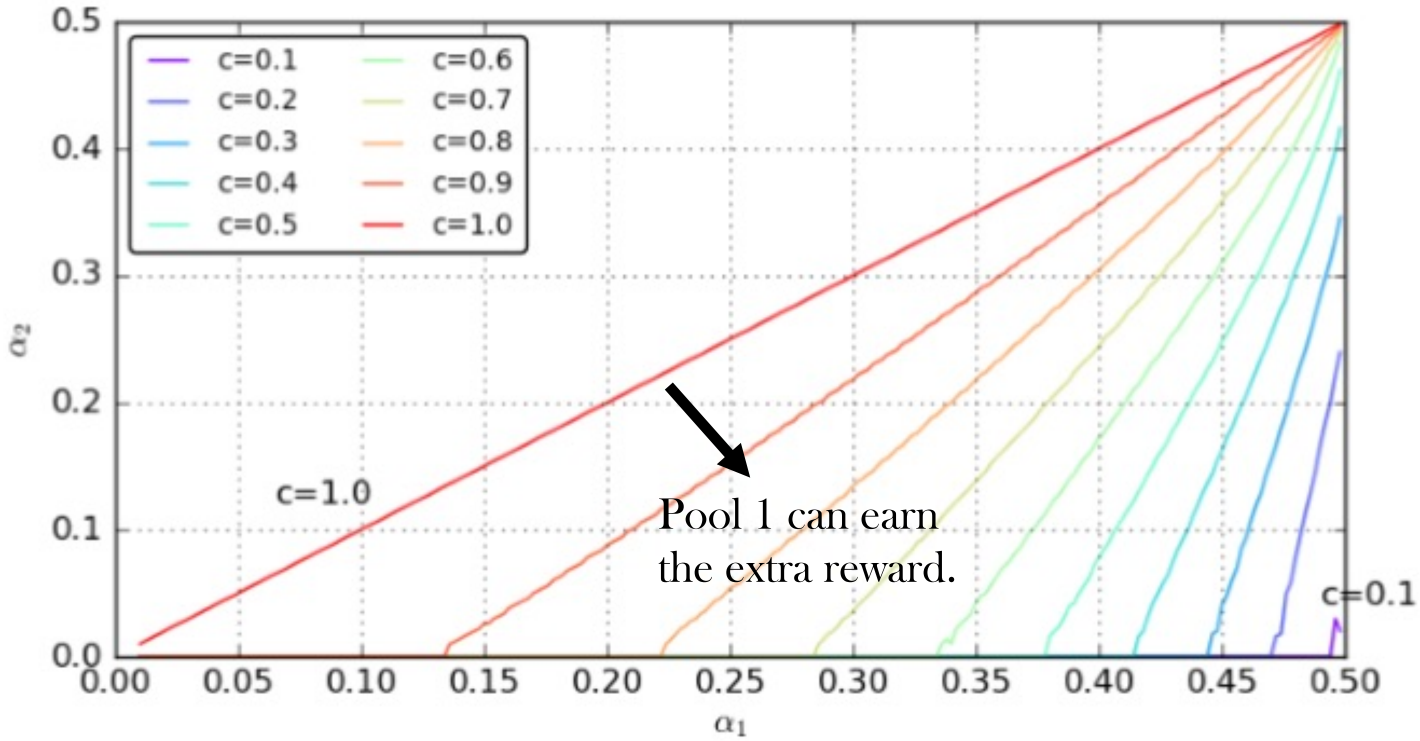


Result



- Pool 1 possesses 0.2 computational power.
- The bigger pool can earn the extra reward unlike the miner's dilemma.

Break Dilemma

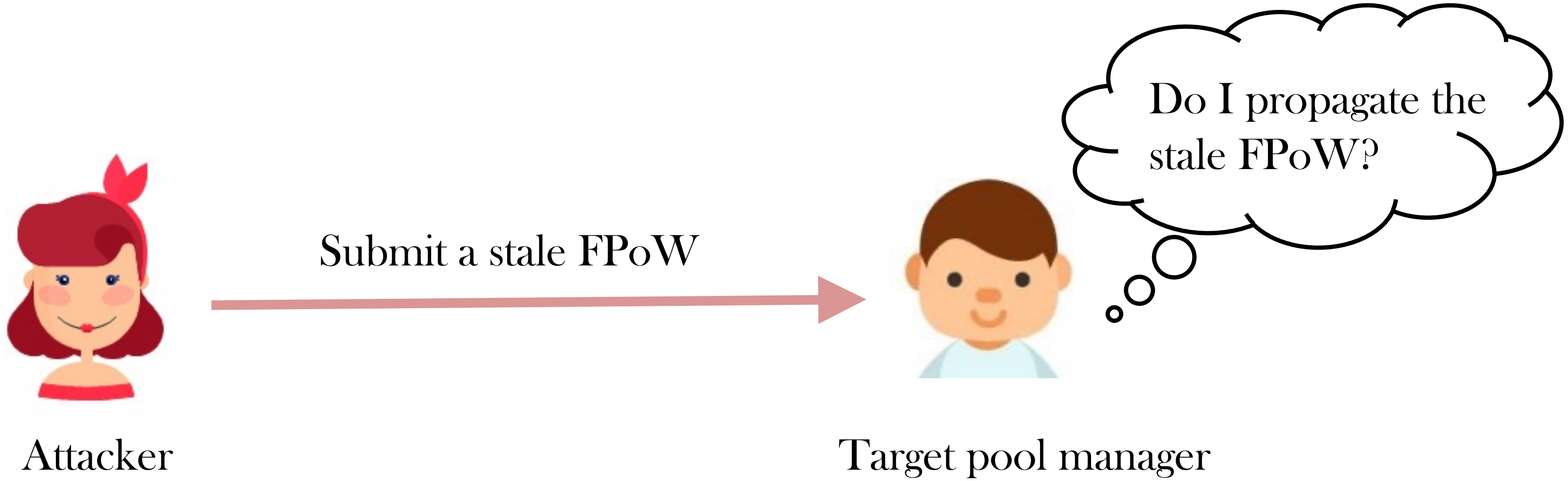


❖ The FAW attack game leads to a pool size game: the larger pool can always earn the extra reward.

FAW Attack VS. Selfish Mining

- The FAW attack is always profitable unlike Selfish mining.
- Selfish miner leave a trace of her identity. However, the FAW attacker leave a trace of the target pools' identity.
  - The rational manager does not propagate immediately blocks which honest miners generate.
  - Forks by selfish mining have unique shape.
- The FAW attack is stealthier than Selfish mining.

Rational Manager



❖ The rational manager should propagate attacker’s FPoWs as fast as possible.

❖ This behavior decreases the manager’s loss and increases the attacker’s reward as a side-effect.

Detection

- ❖ The FAW attack is easier to detect than the BWH attack because of the high fork rate.
- ❖ The manager should suspect and expel any miner who submits stale FPoWs, rather than paying out the reward for the current round.
- ❖ The attacker may easily launch the attack using many Sybil nodes with many churns, replacing the expelled miner.
- ❖ The behavior makes detection useless.

No Silver Bullet

- Detection
  - Beacon value
  - Honeypots
  - An attacker can be rarely affected by the detection.

- New reward system
  - High variance of rewards

- Change Bitcoin protocol
  - Two-phase proof-of-work
  - Not backward compability

- **There is no one silver bullet.**

The FAW Attack is Stronger Than Existing Attacks.

Thank You!

syssec@kaist.ac.kr