采用工作量证明共识机制的区块链中挖矿攻击者间的“鲶鱼效应”

阮 娜[1)] 刘汉卿[1)] 斯雪明[2),3)]

[1)](上海交通大学电子信息与电气工程学院 上海 200240)

[2)](中原工学院前沿信息技术研究院 郑州 450007)

[3)](复旦大学计算机科学技术学院上海市数据科学重点实验室 上海 201203)

摘 要 近年来,采用工作量证明共识机制(Proof of Work,PoW)的区块链被广泛地应用于以比特币为代表的数字加密货币中.自私挖矿攻击(Selfish mining)等挖矿攻击(Mining attack)策略威胁了采用工作量证明共识机制的区块链的安全性.在自私挖矿攻击策略被提出之后,研究者们进一步优化了单个攻击者的挖矿攻击策略.在前人工作的基础上,本文提出了新颖的两阶段挖矿攻击模型,该模型包含拥有单攻击者的传统自私挖矿系统与拥有两个攻击者的多攻击者系统.本文的模型同时提供了理论分析与仿真量化分析,并将两个攻击者区分为内部攻击者与外部攻击者.通过引入内部攻击者与外部攻击者的概念,本文指出传统自私挖矿系统转化为多攻击者系统的条件.本文进一步揭示了在多攻击者系统中两个攻击者将产生竞争并面临着“矿工困境”问题.攻击者间的竞争可被总结为“鲶鱼效应”:外部攻击者的出现导致内部攻击者的相对收益下降至多 67.4%,因此内部攻击者需要优化攻击策略.本文提出了名为部分主动发布策略的全新挖矿攻击策略,相较于自私挖矿策略,该策略是半诚实的攻击策略.在特定场景下,部分主动发布策略可以提高攻击者的相对收益并破解攻击者面临的“矿工困境”问题.

关键词 区块链;比特币;工作量证明共识机制;挖矿攻击;自私挖矿

中图法分类号 TP311 **DOI 号** 10.11897/SP.J.1016.2021.00177

Catfish Effect Between Selfish Miners in Proof-of-Work Based Blockchain

RUAN Na[1)] LIU Han-Qing[1)] SI Xue-Ming[2),3)]

[1)](*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai* 200240)

[2)](*Zhongyuan University of Technology, Zhengzhou* 450007)

[3)](*Shanghai Key Laboratory of Data Science, School of Computer Science and Technology, Fudan University, Shanghai* 201203)

Abstract In the past decade, the consensus mechanism named proof of work has been widely applied by cryptocurrencies like Bitcoin. Though the security of a Proof of Work powered cryptocurrency is always the top priority, it is threatened by mining attacks like selfish mining. In 2014, Eyal proposed the first selfish mining model in a Proof of Work powered blockchain. Selfish mining is an irrational strategy so that the absolute value of the attacker's reward will drop. However the ratio of the attacker's reward to the other miner's block reward would increase. Thus, the relative reward of the attacker is used to measure whether his attacking strategy is effective or not. Currently, the vast computational power in cryptocurrencies like Bitcoin and Ethereum makes it unrealistic to launch a selfish mining attack against them. But mining attack strategies are continuously optimized in recent years. After Eyal's work, researchers have proposed many mining attack models with a single attacker. In these mining attack models,

收稿日期:2019-12-02;在线发布日期:2020-07-28.本课题得到国家自然科学基金青年基金(61702330)资助.**阮 娜**,博士,副教授,中国计算机学会(CCF)专业会员,主要研究方向为安全与隐私、区块链与大数据.E-mail:naruan@cs.sjtu.edu.cn.**刘汉卿**,硕士,主要研究方向为区块链安全.**斯雪明**,教授,中国计算机学会(CCF)会员,主要研究领域为密码学、数据科学、计算机体系结构、网络安全、区块链.

the attacker's strategy space is extended and optimized via Markov decision process. The extension of the attacker's strategy space is based on the assumption that the attacker is the only adversarial node in the blockchain network so that the strategy space of an attacker is still limited. In this paper, we propose a mining attack model with two attackers: the internal attacker and the external attacker. According to the attacker's order of appearance, we divide the model into two phases: the traditional selfish mining system with the internal attacker and the multi-attacker system with the internal attacker and the external attacker. Our model provides both theoretical and quantitative analysis of how the traditional selfish mining system turns into the multi-attacker system. We prove that after the occurrence of the external attacker, both attackers will face a dilemma in the multi-attacker system: Both attackers' relative reward is less than expected. The unexpected competitions between the internal attacker and the external attacker, the overestimation of an attacker's influence rate by himself and the auction like behavior of the internal attacker and the external attacker in the multi-attacker system are three leading causes of two attackers' dilemma. We name this phenomenon as Catfish effect in the multi-attacker system. The Catfish effect in the multi-attacker system leads to the consequence that the internal attacker or the external attacker has to optimize his attacking strategy. The internal attacker would overestimate his relative reward in the multi-attacker system by up to 67.4% after the existence of the external attacker. To break the dilemma faced by the internal attacker and the external attacker, we propose a novel mining strategy named Partial Initiative Release, which is a semi-honest mining strategy. An attacker makes his action based on Partial Initiative Release state machine. The most significant difference between Partial Initiative Release and selfish mining is that the attacker might act as an honest miner when he is at some certain state. In some specific situations, Partial Initiative Release allows the attackers to achieve a higher relative reward compared with selfish mining.

Keywords blockchain; Bitcoin; proof-of-work; mining attack; selfish mining

1 引 言

随着区块链技术的发展,去中心化的数字加密货币逐渐被重视.以比特币[1]为代表的去中心化数字加密货币通过区块链技术保证其安全性.共识机制是区块链技术的核心之一,比特币采用了工作量证明共识机制,该共识机制保证了比特币中交易的一致性与不可篡改性.由于比特币的先发优势,基于工作量证明区块链的数字加密货币占据了大部分市场份额[2].在采用工作量证明共识机制的区块链中,节点以消耗计算力解决特定密码学问题的方式发现新区块,这些节点被称为矿工节点.矿工尝试发现新区块的行为被称为挖矿.当某一矿工发现新区块时,该新区块将被传播至区块链网络中的其他矿工处,其他矿工需要确认该新区块的合法性.若该区块合法,其他矿工将在这个区块继续挖矿.挖到合法的新区块的矿工将收到一定数额的区块奖励和区块中包含交易的手续费作为挖矿收益.采用工作量证明共识机制的区块链中的挖矿难度随着算力不断加入而上升,矿工独自挖矿并最终发现新区块的概率也随之下降.为了稳定地获得挖矿收益,拥有较小算力的矿工自发组织起来并形成矿池[3].

拥有高额算力的矿池或矿工威胁了采用工作量证明共识机制的区块链的安全性.在最理想的情况下,当矿工或矿池拥有的算力在全区块链网络算力中的占比为 σ 时,该矿工或矿池获得的区块奖励在全区块链网络中的占比也为 σ.但是,现有的研究指出,攻击者可以通过自私挖矿攻击[4-6]、区块截留攻击(Block withholding attack)[3]等方式获得额外的挖矿收益.在自私挖矿攻击中,攻击者在区块链中主动制造分叉链(即攻击者的私链),并根据主链高度、攻击者的私链高度等因素选择性地发布区块.在区块截留攻击中,攻击者派遣算力至受攻击矿池,被派遣的算力为受害矿池工作并在挖到全工作量证明(Full Proof of Work,FPoW)区块时抛弃该区块.当

区块截留攻击模型中存在多个攻击者时，攻击者将面临囚徒困境[3,7]，该现象也被描述为“矿工困境”(Miner's dilemma).现有研究也提出了FAW攻击[8](Fork After Withholding attack)、PAW攻击[9](Power Adjusting Withholding)、Selfholding攻击[10]等攻击策略以破解矿工困境.

自私挖矿攻击模型中也可以引入多个攻击者.Liu等研究者[11]率先提出包含多个攻击者的挖矿攻击模型.随后，Katz等研究者[12]与Bai等研究者[13]进一步研究了多攻击者挖矿攻击模型.但是上述研究者没有进一步分析多攻击者挖矿攻击模型中出现多个攻击者的原因.同时，在多攻击者挖矿攻击模型中，攻击者同样面临“矿工困境”，上述研究者没有详细描述该问题.

本文通过理论分析与仿真量化分析解释了攻击模型中出现两个攻击者的原因，并提出了全新的两阶段挖矿攻击模型.本文第3节提出该挖矿攻击模型的两个阶段：传统自私挖矿系统与多攻击者系统；第4节引入内部攻击者与外部攻击者的概念，并解释了攻击模型从传统自私挖矿系统向多攻击者系统转化的条件；第5节解释了两个攻击者面临“矿工困境”的原因；第6节揭示了攻击者间的“鲶鱼效应”并提出了全新的半诚实攻击策略：部分主动发布策略，该半诚实的攻击策略可以打破攻击者面临的“矿工困境”.

本文的主要贡献可以归纳为以下三点：

(1)建立了两阶段挖矿攻击模型并通过理论分析与仿真量化分析解释模型中出现两个攻击者的原因：提出了全新的两阶段挖矿攻击模型，并引入了内部攻击者与外部攻击者的概念.本文指出内部攻击者将导致传统自私挖矿系统有效算力的下降，并首次提出收缩系数的概念以度量内部攻击者对传统自私挖矿系统的影响力.本文还证明了当外部攻击者算力满足一定条件时，外部攻击者将攻击传统自私挖矿系统.

(2)提出了多攻击者挖矿攻击模型中攻击者面临“矿工困境”：攻击者预期之外的竞争、攻击者拍卖式发布区块与攻击者高估自身影响力导致两个攻击者收益下降，从而使攻击者面临“矿工困境”.

(3)揭示攻击者间的鲶鱼效应并提出全新的挖矿攻击策略：讨论了两阶段挖矿攻击模型多攻击者系统中的“鲶鱼效应”，并提出了名为部分主动发布区块策略的全新的挖矿攻击策略.两阶段挖矿攻击模型中的攻击者可以使用该策略提升自己的相对收益并避免“矿工困境”问题.

2　相关工作

2.1　比特币基础

化名为中本聪的学者在提出区块链的概念时，同时提出了比特币[1].因此，比特币是最具有代表性的基于区块链技术的数字加密货币.比特币之后的大部分数字加密货币均参考了比特币的架构.本节通过介绍比特币中的要素展现基于工作量证明区块链的数字加密货币的特性.

比特币中的区块由区块头与区块体构成.区块头区分了比特币中的每一个区块，一个区块的区块头包含了上一区块区块头的哈希值、区块中所有交易的梅克尔根(Merkle root)[14]与一个随机数.一个区块的区块体存储了被该区块确认的交易信息.矿工的工作是选取尚未被确认的交易以及生成一个随机数.当矿工选取的交易与生成的随机数使得区块头的哈希值拥有一定数量的前导零时，矿工成功挖到新区块.该新区块将被矿工传播至区块链网络.其余矿工在确认新区块的合法性后，将接受该区块并将该区块加入主链的末端.矿工挖矿的过程消耗了计算力，因此，比特币采用的共识机制被称为工作量证明共识机制.比特币通过规定新区块区块头哈希值前导零数量的方式调整挖矿难度，比特币中的算力越多，挖矿难度越大.通常来说，比特币将新区块产生的速度控制为每10 min产生一个新区块.比特币的挖矿难度约两周调整一次.

比特币的价格激励了大量计算力加入比特币系统并挖矿.随着挖矿难度的上升，对于拥有较小算力的独立矿工来说，短期内发现新区块并获得挖矿奖励的概率几乎为零[15].因此，为了获得更加稳定的挖矿收益，独立矿工们被组织起来，形成了矿池.比特币历史上曾出现过算力在全比特币系统中占比超过40%的矿池[2].矿池在为其成员带来了更加稳定的收益的同时，威胁到了比特币的安全性.

比特币网络中产生的分叉(fork)可被分为两类：自然产生的分叉与攻击者制造的分叉.自然产生的分叉通常由诚实矿工在不经意间制造，由于比特币网络中存在网络传播时延，两个诚实矿工可能在接近同一时刻发布区块.Gervais等研究者[2]与Decker等研究者[16]对区块链网络中自然产生分叉的概率做了预测.他们的研究显示，比特币网络中产生自然分叉的概率处于0.41%至1.7%之间.攻击

者制造的分叉由攻击者蓄意制造，在自私挖矿攻击等攻击策略中，攻击者通过蓄意制造分叉链的方式在区块链网络中发起竞争.

在比特币中，遵守比特币协议挖矿的矿工被称为诚实矿工. 在发现新区块后，诚实矿工将立即向区块链网络广播. 同时，诚实矿工将区块链网络中的最长链视作主链. 当区块链网络中出现高度相同的分叉链时，诚实矿工将最先接收到的分叉链视作主链. 比特币中攻击者的行为与诚实矿工不同. 一些典型的攻击者行为包括：

(1) 攻击者拒绝向区块链网络传播其他矿工挖出的区块[17-20].

(2) 在攻击者发现新区块后，攻击者拒绝立即向区块链网络传播区块信息[4-6,8].

(3) 攻击者拒绝接受区块链网络中的最长链作为主链[4,17,21-22].

2.2 自私挖矿攻击

Eyal 等研究者率先提出自私挖矿[4]. 自私挖矿攻击的基本思想是：攻击者制造分叉链且选择时机发布分叉链中的区块. Eyal 等研究者将攻击者的状态定义为攻击者的分叉链领先主链的长度并根据攻击者的状态、状态转移概率给出攻击者的状态机. 同时，Eyal 为攻击者设计了四个基本动作：放弃(Adopt)、匹配(Match)、持有(Hold)与覆盖(Override). 当且仅当攻击者处在特定状态时，攻击者会主动发布区块(匹配或覆盖).

图 1 展示了自私挖矿攻击中的三个典型状态. 通过图 1 中的三个状态可以很好地总结自私挖矿攻击的特点. 状态一展示了自私挖矿攻击者的 Hold 操作：当诚实矿工未发布区块时，自私挖矿攻击者执行 Hold 操作不发布区块；状态二展示了自私挖矿攻击者的 Override 操作：当诚实矿工挖出区块且自私挖矿攻击者的分叉链长度为二时，攻击者将发布整个分叉链，发布后的分叉链将成为新的主链；状态三展示了自私挖矿攻击者的 Match 操作：当诚实矿工挖出区块且自私挖矿攻击者的分叉链长度为一时，攻击者将发布整个分叉链，此时诚实矿工与自私挖矿攻击者间产生竞争. 由此可见，自私挖矿攻击者仅在特定的状态发布，这是自私挖矿攻击最显著的特点.

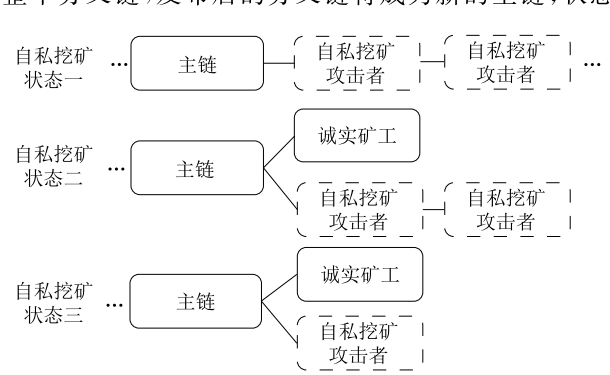


图 1 自私挖矿攻击典型状态(虚线代表攻击者未发布的区块)

自私挖矿攻击策略大幅降低了攻击者对区块链系统发动攻击所需要的算力. 攻击者算力在区块链系统中占比为 1/3 时，攻击者即可确保其通过攻击获得额外收益.

在自私挖矿攻击被提出后，Sapirshtein 等研究者[6]与 Nayak 等研究者[5]拓展了自私挖矿攻击者的策略集. Sapirshtein 等研究者利用马尔科夫决策，提出了优化自私挖矿策略，Nayak 等研究者将自私挖矿攻击与区块链网络层攻击日蚀攻击结合，提出了固执挖矿攻击. Liu 等研究者[11]探讨了自私挖矿模型中出现多个攻击者时的情况.

自私挖矿攻击同时浪费了攻击者的算力与诚实矿工的算力，是短期内非理性的攻击策略[2]. 因此，攻击者获得区块奖励的绝对数额无法直接衡量挖矿攻击策略的表现. 但是，挖矿攻击可以提升攻击者的相对收益，即攻击者的收益占比. 从长期来看，随着比特币挖矿难度的调整，攻击者的相对收益将转换为攻击者的绝对收益. 通常来说，研究者们使用相对收益(Relative Reward，RR)与孤块率(Stale Block Rate，SBR)比较挖矿攻击策略的表现.

2.3 攻击者间的博弈问题

Eyal 等研究者首先提出区块截留攻击模型中攻击者面临“囚徒困境”[3]：当多个矿池同时采用区块截留攻击策略时，所有的攻击者将遭受损失. Eyal 将该现象命名为“矿工困境”. 唐长兵等研究者[7]利用零行列式策略对区块截留攻击中的矿工策略选择进行优化. 此后，Kwon 等研究者[8]、Gao 等研究者[9]与 Dong 等研究者[10]分别提出 FAW 攻击、PAW 攻击与 Selfholding 攻击帮助攻击者避免“矿工困境”问题.

在挖矿攻击模型中引入多个攻击者时，攻击者同样面临“矿工困境”问题. 但是，目前挖矿攻击模型中多个攻击者间的博弈问题尚未被深入研究. 本文作者[11]率先提出了包含多个攻击者的挖矿攻击模型. 但是本文作者的前期工作仅通过数值仿真分析了多个攻击者的收益情况，并未进一步分析攻击者间的博弈问题. Katz 等研究者[12]的工作是与本文同步开展的工作，Katz 等研究者同样提出了一种半诚

实的挖矿攻击策略.本文的工作与 Katz 等研究者工作的区别在于:(1)半诚实挖矿攻击策略的区别;(2)本文指出了挖矿攻击模型中攻击者们面临的“矿工困境”问题,并将半诚实挖矿攻击策略用于破解“矿工困境”问题.

3　攻击模型

3.1　两阶段挖矿攻击模型

本文提出的两阶段挖矿攻击模型由两个部分构成:传统自私挖矿系统与多攻击者系统.如图 2 所示,两阶段挖矿攻击模型的第一阶段为传统自私挖矿阶段.在传统自私挖矿阶段,内部攻击者针对诚实矿工实施自私挖矿攻击.内部攻击者与诚实矿工组成了传统自私挖矿系统.在内部攻击者实施发动攻击后.传统自私挖矿系统的区块链出现分叉,区块链分叉的后果是传统自私挖矿系统的有效算力下降.有效算力下降的传统自私挖矿系统易被外部攻击者视作目标.外部攻击者将针对传统自私挖矿系统发动挖矿攻击.此时,两阶段挖矿攻击模型进入多攻击者阶段.内部攻击者、诚实矿工与外部攻击者组成了多攻击者系统.

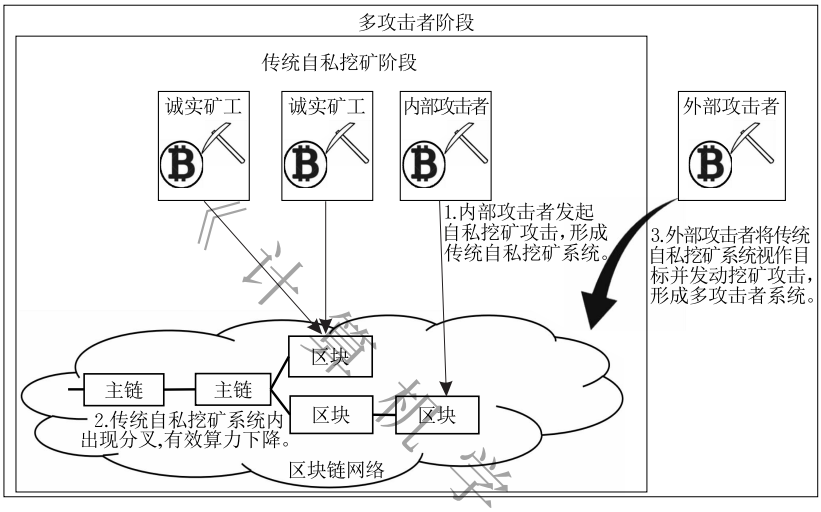


图 2　两阶段挖矿攻击模型

3.2　攻击者状态与行为

两阶段挖矿攻击模型中,对攻击者状态与攻击者行为的描述与前人的工作保持一致.

攻击者状态.攻击者的状态应包含如下信息:攻击者未发布的私有链领先诚实矿工的区块数与攻击者是否有参与到分叉链的竞争中.本文的模型使用标记 $S=i(i=0,1,2\cdots)$ 表示攻击者的私有链领先诚实矿工的主链 i 个区块,且攻击者没有参与到区块链的竞争中.$S=0'$ 表示攻击者正在与其他矿工竞争.$S=i'(i=1,2,3\cdots)$ 表示攻击者的私有链领先诚实矿工的主链 i 个区块,且区块链网络中存在着攻击者并未参与的竞争.

攻击者行为.攻击者的行为应包含如下信息:攻击者是否发布区块与攻击者发布区块的数量.攻击者在区块链网络中的所有动作可以用以下五个行为描述:(1)Hold:攻击者不发布任何区块;(2)Match:攻击者发布至少一个区块,从而在区块链网络中制造一个与当前主链长度相同的分叉链;(3)Override:攻击者发布至少两个区块,从而在区块链网络中制造一个比当前主链长度多一个区块的分叉链;(4)Adopt:攻击者放弃自己当前的私链;(5)Release:攻击者主动发布一个区块.

3.3　攻击者的表现

攻击者的孤块率与攻击者相对收益可以用来衡量攻击者的表现.

攻击者的孤块率体现了攻击者发起挖矿攻击的代价.将攻击者挖到并被全区块链网络承认的区块数量记作 N_{ac},将攻击者挖到却并未被全区块链网络接收的区块数量记作 N_{nac},将攻击者的孤块率记作 SBR:

$$SBR=\frac{N_{nac}}{N_{nac}+N_{ac}} \tag{1}$$

攻击者的相对收益体现了攻击者获得的收益在全区块链网络中的占比.当攻击者的相对收益大于攻击者的计算力在全区块链网络中的占比时,攻击者的攻击行为是成功的.将除攻击者以外所有矿工挖到并被全区块链网络承认的区块数量记作 N_{oac},将攻击者的相对收益记作 RR:

$$RR=\frac{N_{ac}}{N_{ac}+N_{oac}} \tag{2}$$

3.4 理想化模型的假设

本文提出的两阶段挖矿攻击模型基于如下假设：

(1)诚实矿工、内部攻击者与外部攻击者是三个不同的实体，三个实体间无交流.

(2)传统自私挖矿系统中只包含内部攻击者，内部攻击者算力小于诚实矿工且内部攻击者不能预测外部攻击者是否存在；在外部攻击者攻击传统自私挖矿系统前，外部攻击者无法感知内部攻击者的存在.

(3)外部攻击者对传统自私挖矿系统发起挖矿攻击前，将传统自私挖矿系统内所有算力都判定为诚实矿工.

(4)为方便计算，多攻击者系统中的全部算力被标准化为1.

(5)传统自私挖矿系统与多攻击者系统不考虑出现自然产生的分叉.

3.5 模型中的参数

两阶段挖矿攻击模型使用了如下参数：

(1)α，β_1与β_2. α，β_1与β_2分别代表诚实矿工、内部攻击者与外部攻击者在多攻击者系统中的算力占比. $\alpha+\beta_1+\beta_2=1$.

(2)α'与β'. α'与β'分别代表诚实矿工与内部攻击者在传统自私挖矿系统内的算力占比. $\alpha'=\frac{\alpha}{\alpha+\beta_1}$，$\beta'=\frac{\beta_1}{\alpha+\beta_1}$.

(3)α_t与β''. α_t代表了外部攻击者在搜寻攻击目标阶段对目标系统算力的估测值. 外部攻击者根据目标系统区块链主链增长速度与挖矿难度估算目标系统算力. β''同样由外部攻击者在搜寻攻击目标阶段计算得出，代表外部攻击者对发起攻击后算力占比的预测值，$\beta''=\frac{\beta_2}{\beta_2+\alpha_t}$.

(4)γ_h. γ_h表示在区块链网络中出现分叉链时，诚实矿工支持攻击者(内部攻击者或外部攻击者)的算力占诚实矿工总算力之比.

(5)γ_1与γ_2. γ_1与γ_2分别表示在区块链网络中出现分叉链时，诚实矿工支持内部攻击者与外部攻击者的算力占诚实矿工总算力之比. γ_1与γ_2分别体现了两个攻击者的影响力.

3.6 模型参数范围

参数α，β_1与β_2的范围难以通过现实情况确定，因为目前尚未出现多个挖矿攻击者同时发起攻击的案例. 但是参数α'与β'仅涉及包含一位攻击者的传统自私挖矿系统，其范围可以通过现实中矿池的规模做出合理假设. 本文通过前人研究[4]与比特币中矿池算力占比确定参数α'与β'的范围. $\alpha'=0.75$且$\beta'=0.25$被研究者们认为是内部攻击者在传统自私挖矿系统内成功发起自私挖矿攻击所需算力的最小值；$\beta'=0.45$近似于比特币历史上出现的最大矿池算力占比，此时$\alpha'=0.55$. 参数α'与β'决定了参数α与参数β_1之比. 参数α_t由外部攻击者估算得出，α_t与参数α、β_1、α'与β'相关. 依据参数α_t可以对参数β_2的范围做出合理的假设.

参数γ_1、γ_2与γ_h的范围可以根据现有研究与实际情况确定. Nayak等研究者[5]指出攻击者影响力的范围可大至$[0,0.92]$(即γ_h的范围是$[0.08,1]$). 在实际情况中，攻击者节点的部署情况、诚实矿工节点的部署情况、节点的动态特性等因素都会导致攻击者影响力的变化. 因此，为了涵盖所有场景，本文将参数γ_h的范围设定为$[0,1]$. 内部攻击者与外部攻击者将划分被影响的诚实矿工算力.

4 内部攻击者与外部攻击者

4.1 传统自私挖矿系统与内部攻击者

两阶段挖矿攻击模型的传统自私挖矿系统包含了内部攻击者与诚实矿工，内部攻击者对传统自私挖矿系统内的诚实矿工发起自私挖矿攻击.

如图3所示的五个状态包含了传统自私挖矿系统区块链网络的所有状态. 这五个状态分别是：内部

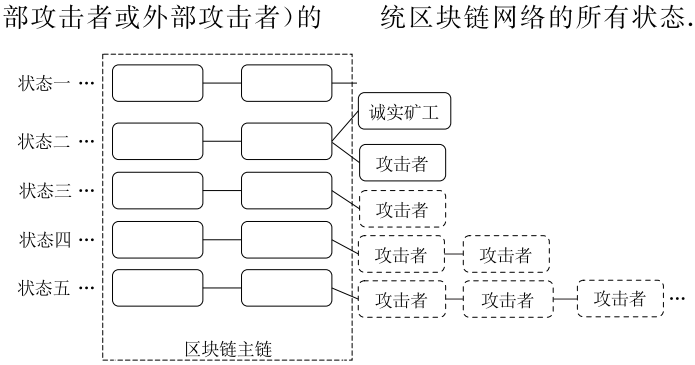


图3 传统自私挖矿系统区块链网络的五个状态

攻击者不持有未发布的私有链、内部攻击者与诚实矿工产生竞争、内部攻击者私有链领先诚实矿工的主链一个区块、内部攻击者私有链领先诚实矿工的主链两个区块与内部攻击者私有链领先诚实矿工的主链多于两个区块.

从现有的研究工作[4]可以推出区链网络处于各状态的概率. 令 α' 与 β' 分别代表诚实矿工与内部攻击者在传统自私挖矿系统内的算力占比. 传统自私挖矿系统处于状态一的概率是 $\frac{\beta'-2\beta'^2}{\beta'(2\beta'^3-4\beta'^2+1)}$；处于状态二的概率是 $\frac{\alpha'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$；处于状态三的概率是 $\frac{\beta'-2\beta'^2}{2\beta'^3-4\beta'^2+1}$；处于状态四的概率是 $\frac{\beta'(\beta'-2\beta'^2)}{\alpha'(2\beta'^3-4\beta'^2+1)}$；处于状态五的概率是 $\frac{\beta'^2(\beta'-2\beta'^2)}{\alpha'(\alpha'-\beta')(2\beta'^3-4\beta'^2+1)}$.

4.2　外部攻击者出现的原因

引理 1.　传统自私挖矿系统中，每当内部攻击者或诚实矿工挖到一个新区块时，主链的长度有一定概率增加，该概率小于 1 且与内部攻击者的影响力 γ_1 无关.

证明.　在传统自私挖矿系统中，诚实矿工有 α' 的几率发现新区块. 无论在诚实矿工发现新区块前，区块链处于何种状态，诚实矿工发现的新区块都将导致主链长度增加. 当区块链网络处于状态一、状态二、状态三与状态五时，诚实矿工能使区块链主链长度增加一个区块；当区块链网络处于状态四时，诚实矿工能使区块链主链长度增加两个区块. 因此，诚实矿工使区块链网络长度增加的期望值为 $\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}$.

内部攻击者有 β' 的概率发现新区块. 在内部攻击者发现新区块前，若区块链网络处于状态一、状态三、状态四与状态五，攻击者挖到的新区块无法使区块链网络延长. 若区块链网络处于状态二，则内部攻击者能使区块链主链长度增加一个区块. 因此，内部攻击者使区块链网络主链长度增加的期望值为 $\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$.

因此，每当一个新区块出现时，传统自私挖矿系统主链增长：$\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$. 主链的增长与内部攻击者的影响力 γ_1 无关. 此时，记 $f(\beta')=\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}-1$，则 $\frac{df(\beta')}{d\beta'}=\frac{-8\beta'^4+9\beta'^2-4\beta'}{4\beta'^6-16\beta'^5-16\beta'^4+4\beta'^3-8\beta'^2+1}$.

当 $0<\beta'<\frac{1}{2}$ 时，$\frac{df(\beta')}{d\beta'}\leq 0$. 而 $f(0)=0$，从而得出 $f(\beta')<0$. 因此当新区块被发现时主链增加的长度小于 1.　　证毕.

区块链系统中的算力可被划分为有效算力与无效算力. 当一区块未被纳入主链成为孤块时，花费在该区块上的算力即被视作无效算力；反之，当区块被纳入主链时，花费在该区块上的算力则为有效算力. 因此，区块链系统的有效算力与挖出一个区块的难度与区块链主链增长速度相关. 当区块链系统中所有节点均为诚实节点时，矿工挖出的每一个区块都可以使区块链网络主链长度增加一个区块. 此时区块链系统内的有效算力等价于区块链系统中所有节点的算力之和. 但是，由引理 1 可知，内部攻击者在传统自私挖矿系统中发起自私挖矿攻击后，每个区块仅有概率 $\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$ 使主链长度增加一个区块，这意味着每个区块有 $1-\alpha'-\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}-\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$ 的概率成为孤块并被废弃，被遗弃的孤块无法增加区块链系统的主链长度. 尽管传统自私挖矿系统中的总算力为 $(\alpha+\beta_1)$，但是根据传统自私挖矿系统主链增长速度与挖矿难度计算得出的系统内有效算力将因为内部攻击者的行为而下降至 $(\alpha+\beta_1)\cdot\left(\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}\right)$. 本文定义收缩系数 $\rho=\alpha'+\frac{(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$，收缩系数可表示出内部攻击者对传统自私挖矿系统造成的影响. 收缩系数 ρ 与传统自私挖矿系统内的总算力 $(\alpha+\beta_1)$ 的乘积 $\rho(\alpha+\beta_1)$ 表示了传统自私挖矿系统的有效算力.

本文通过蒙特卡洛方法模拟区块的生成，并生成了主链长度为 10^6 的区块链.

图 4 展示了仿真中 γ_1 取不同值时，传统自私挖矿系统收缩系数与内部攻击者算力在传统自私挖矿系统中的占比的关系. 仿真结果说明，内部攻击者的

影响力 γ_1 的取值不影响传统自私挖矿系统的收缩系数.这与引理 1 相符.此外,攻击者算力在传统自私挖矿系统中占比越高,传统自私挖矿系统的收缩系数越小.这表明,攻击者算力越高,传统自私挖矿系统内的有效算力越低.

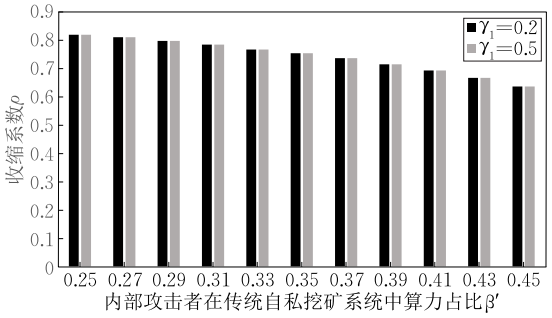


图 4 传统自私挖矿系统中的收缩系数

本文的理论分析说明,收缩系数 $\rho=\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$.表 1 展示了收缩系数 ρ 在理论中与仿真中的对比.表 1 中的误差率说明了传统自私挖矿系统的收缩系数 ρ 是可预测的,内部攻击者在传统自私挖矿系统中发起自私挖矿攻击前即可预测到传统自私挖矿系统有效算力下降的程度.

表 1 传统自私挖矿系统收缩系数 ρ 在理论中与仿真中的对比

β'	ρ 理论值	ρ 仿真值	误差值/%
0.25	0.81995	0.82000	0.0059
0.29	0.79482	0.79479	−0.0046
0.33	0.76714	0.76718	0.0053
0.37	0.73471	0.73478	0.0098
0.41	0.69363	0.69336	−0.0381
0.45	0.63485	0.63431	−0.0843

4.3 外部攻击者出现的条件

在外部攻击者寻找攻击目标时,外部攻击者将根据目标系统的挖矿难度与其主链增长的速度推测目标系统的算力,本文将外部攻击者推测的目标系统算力记作 α_t.在外部攻击者的推测中,外部攻击者(具有算力 β_2)与目标系统(具有算力 α_t)将构成一个新的传统自私挖矿系统,该系统具有算力 $\beta_2+\alpha_t$.外部攻击者在该系统内的算力占比 $\beta''=\frac{\beta_2}{\beta_2+\alpha_t}$.

引理 2. 若外部攻击者算力满足 $\beta_2\geq\frac{\alpha_t(1-\gamma_2)}{2-\gamma_2}$,外部攻击者将对目标系统发起自私挖矿攻击,其中 γ_2 为外部攻击者在目标系统内的影响力.

证明. 将外部攻击者的相对收益记作 RR_{ex},则外部攻击者的目标为 $RR_{ex}\geq\beta''$;将外部攻击者在与目标系统中矿工的竞争中获胜的概率记作 τ_2,$\tau_2=\beta''+\gamma_2\alpha''$.由于外部攻击者将目标系统视作诚实矿工,因此在外部攻击者的预期中,外部攻击者与目标系统构成了传统自私挖矿系统.外部攻击者处于传统自私挖矿系统状态一的概率为 $\frac{\beta''-2\beta''^2}{\beta''(2\beta''^3-4\beta''^2+1)}$;外部攻击者处于状态二的概率为 $\frac{(1-\beta'')(\beta''-2\beta''^2)}{2\beta''^3-4\beta''^2+1}$;外部攻击者处于状态三的概率为 $\frac{\beta''-2\beta''^2}{2\beta''^3-4\beta''^2+1}$;外部攻击者处于状态四的概率为 $\frac{\beta''(\beta''-2\beta''^2)}{(1-\beta'')(2\beta''^3-4\beta''^2+1)}$;外部攻击者处于状态五的概率为 $\frac{\beta''^3}{(1-\beta'')(2\beta''^3-4\beta''^2+1)}$.

当外部攻击者处于状态一时,外部攻击者的收益为 0.当外部攻击者处于状态二时,外部攻击者挖出下一个区块并赢得竞争的概率为 β'',此时收益为 2;受外部攻击者影响的诚实矿工挖出下一个区块并导致外部攻击者在竞争中获胜的概率为 $\tau_2-\beta''$,外部攻击者收益为 1.当外部攻击者处于状态三时,无论外部攻击者是否挖到下一个区块,外部攻击者仅会有状态的转移,而没有直接收益,因此收益为 0.当外部攻击者处于状态四时,若外部攻击者挖到下一个区块,外部攻击者仅会有状态的转移,若诚实矿工挖到区块(概率为 $1-\beta''$),外部攻击者将会直接发布两个区块,因此收益为 2.外部攻击者处于状态五时,无论外部攻击者是否挖到下一个区块,外部攻击者仅会有状态的转移,而没有直接收益,因此收益为 0.

因此在外部攻击者的预测中,外部攻击者的期望收益 E_{ex} 可以表示为

$$E_{ex}=\frac{2\beta''(1-\beta'')(\beta''-2\beta''^2)}{2\beta''^3-4\beta''^2+1}+\frac{(\tau_2-\beta'')(1-\beta'')(\beta''-2\beta''^2)}{2\beta''^3-4\beta''^2+1}+\frac{2\beta''(\beta''-2\beta''^2)}{2\beta''^3-4\beta''^2+1}+\frac{\beta''^3}{(1-\beta'')(2\beta''^3-4\beta''^2+1)} \tag{3}$$

目标系统中的矿工收益来自于:(1)当外部攻击者处于状态二时,受外部攻击者影响的诚实矿工挖到下一个区块;(2)当外部攻击者处于状态二时,未受外部攻击者影响的诚实矿工挖到下一个区块;(3)当外部攻击者处于状态一时,诚实矿工挖到下

一个区块.

目标系统中的矿工的期望收益 E_h 可以表示为

$$E_h=\frac{(\tau_2-\beta'')(1-\beta'')(\beta''-2\beta''^2)}{2\beta''^3-4\beta''^2+1}+\frac{2(1-\tau_2)(1-\beta'')(\beta''-2\beta''^2)}{2\beta''^3-4\beta''^2+1}+\frac{(1-\beta'')(\beta''-2\beta''^2)}{\beta''(2\beta''^3-4\beta''^2+1)} \tag{4}$$

依据目标系统中的矿工与外部攻击者的期望收益，可得外部攻击者相对收益：

$$RR_{ex}=\frac{E_{ex}}{E_{ex}+E_h}\geq\beta'' \tag{5}$$

由不等式(5)及 $\beta''=\frac{\beta_2}{\beta_2+\alpha_t}$ 可推得：

$$(1-\tau_2)(\beta_2^2-\alpha_t^2)+\alpha_t\beta_2\geq 0 \tag{6}$$

将 τ_2 代入后可推出外部攻击者算力 β_2 与目标系统算力 α_t 的关系：

$$\beta_2\geq\frac{\alpha_t(1-\gamma_2)}{2-\gamma_2} \tag{7}$$

证毕.

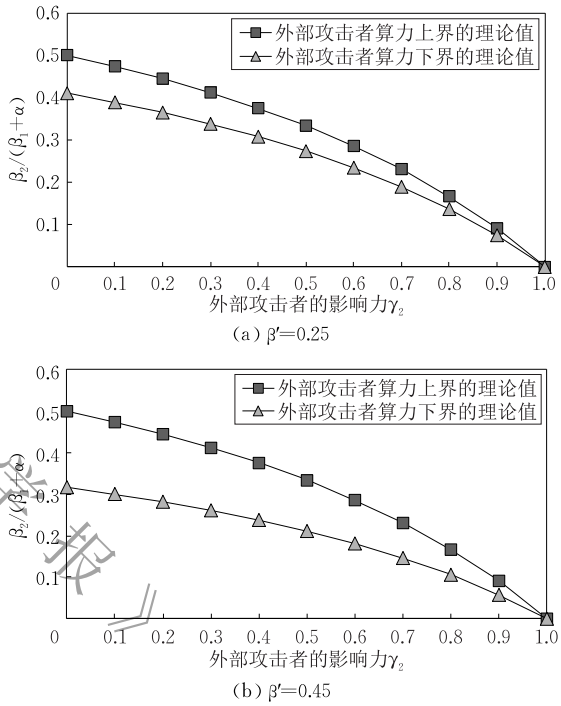
引理 3.　传统自私挖矿系统中，内部攻击者算力为 β_1，诚实矿工算力为 α，收缩系数 $\rho=\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$. 当且仅当外部攻击者算力满足：$\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)\leq\beta_2\leq\frac{1-\gamma_2}{2-\gamma_2}(\alpha+\beta_1)$ 时，外部攻击者会在内部攻击者发起攻击后对传统自私挖矿系统发起自私挖矿攻击.

证明.　从外部攻击者的视角看，由于内部攻击者已在传统自私挖矿系统内发起自私挖矿攻击，传统自私挖矿系统的有效算力为 $\rho(\alpha+\beta_1)$. 由引理 3 可知，外部攻击者对传统自私挖矿系统发起攻击的条件是：$\beta_2\geq\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)$. 与此同时，若 $\beta_2>\frac{1-\gamma_2}{2-\gamma_2}(\alpha+\beta_1)$，外部攻击者在内部攻击者发起攻击之前便将传统自私挖矿系统视作目标. 因此，$\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)\leq\beta_2\leq\frac{1-\gamma_2}{2-\gamma_2}(\alpha+\beta_1)$.　证毕.

$\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)$ 与 $\frac{1-\gamma_2}{2-\gamma_2}(\alpha+\beta_1)$ 可分别被视作外部攻击者算力的下界与上界. 外部攻击者算力的下界与上界受外部攻击者的影响力、内部攻击者的算力与诚实矿工的算力影响. 当外部攻击者算力低于下界时，外部攻击者无法通过发起自私挖矿攻击获得额外收益；当外部攻击者算力高于上界时，外部攻击者无需等待内部攻击者出现即可发起自私挖矿攻击.

在仿真中，本文通过蒙特卡洛方法模拟区块的生成，并生成了主链长度为 10^6 的区块链.

图 5(a)与图 5(b)分别展示了 β' 取值 0.25 与 0.45 时，外部攻击者视角下外部攻击者发起攻击所需要的算力的上界与下界. 由图 5 可以看出，β' 的增长不影响外部攻击者发起攻击所需算力的上界，但是外部攻击者发起攻击所需算力的下界将随 β' 的增长而降低.



(a) $\beta'=0.25$

(b) $\beta'=0.45$

图 5　外部攻击者预测中所需算力上界与下界

5　多攻击者系统中的“矿工困境”

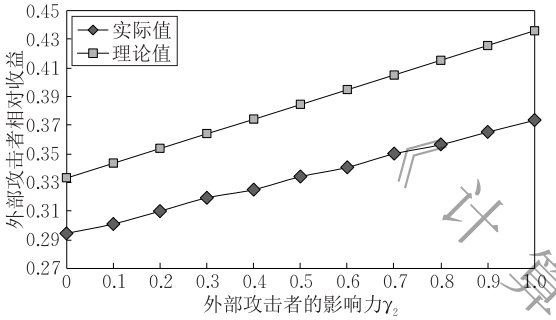
在外部攻击者出现后，两阶段挖矿攻击模型由传统自私挖矿系统转变为多攻击者系统. 在多攻击者系统中，外部攻击者与内部攻击者都将错误地预估自己的攻击收益. 采用自私挖矿攻击策略的内部攻击者与外部攻击者将面临“矿工困境”问题：内部攻击者与外部攻击者同时发动攻击导致了两个攻击者的收益同时下降.

5.1　外部攻击者的困境

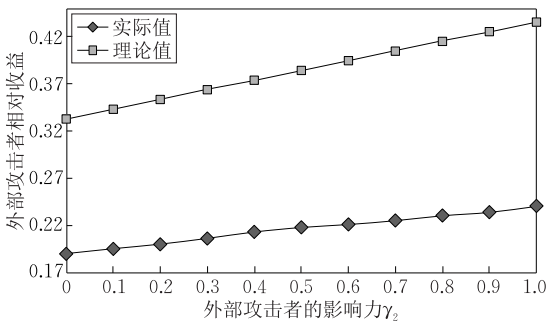
在外部攻击者的视角下，一旦外部攻击者的算力高于阈值 $\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)$，外部攻击者便可通过发

起自私挖矿攻击获取额外收益.然而,该预期是不契合实际的.

通过仿真可以说明外部攻击者往往高估自身的攻击收益.在仿真中,β'的取值仍然为 0.25 与 0.45,对应的收缩系数 ρ 的取值则分别为 0.82 与 0.634.同时,令外部攻击者算力取值 $\frac{1-\gamma_2}{2-\gamma_2}\times\frac{1+\rho}{2}\times(\alpha+\beta_1)$,该算力确保外部攻击者的算力始终处于上界与下界之间.图 6(a)与图 6(b)分别展示了外部攻击者在 $\beta'=0.25$ 与 $\beta'=0.45$ 时,获得相对收益的理论值与实际值.仿真结果表明,外部攻击者在任何 β' 取值下都将高估自身的相对收益.β'取值越高,外部攻击者预估误差越大.



(a) $\beta'=0.25$



(b) $\beta'=0.45$

图 6 外部攻击者相对收益的理论值与预测值

表 2 与表 3 说明了外部攻击者如何陷入“矿工困境”.当 $\beta'=0.45$ 时,外部攻击者估算的误差可以高达 46.66%.这是由于随着 β'值的增加,传统自私挖矿系统收缩系数降低,传统自私挖矿系统的有效算力降低,这导致外部攻击者低估了发动自私挖矿攻击所需要的算力.

表 2 $\beta'=0.25$ 时外部攻击者的相对收益

γ_2	理论值	实际值	误差值/%
0	0.333	0.294	−11.68
0.2	0.354	0.310	−12.40
0.4	0.374	0.325	−13.17
0.6	0.395	0.341	−13.73
0.8	0.415	0.357	−14.16
1.0	0.436	0.374	−14.30

表 3 $\beta'=0.45$ 时外部攻击者的相对收益

γ_2	理论值	实际值	误差值/%
0	0.333	0.190	−42.89
0.2	0.354	0.201	−43.31
0.4	0.374	0.214	−42.93
0.6	0.395	0.222	−43.87
0.8	0.415	0.231	−44.37
1.0	0.436	0.241	−44.66

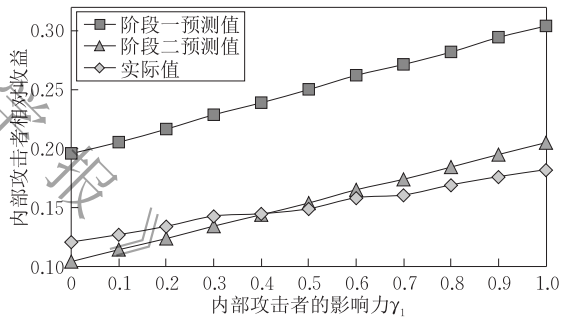
5.2 内部攻击者的困境

由于外部攻击者的出现,内部攻击者将分两个阶段调整对自身相对收益的预期.

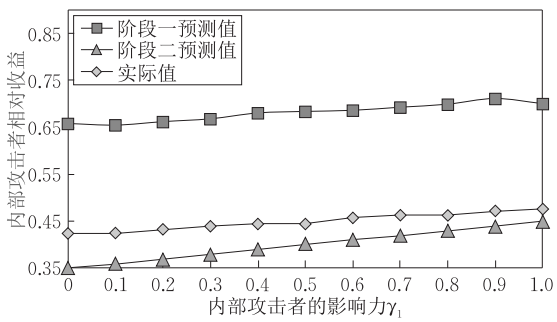
(1)阶段一.内部攻击者尚未发现外部算力的加入.此时,内部攻击者基于传统自私挖矿系统的参数预测自身相对收益.

(2)阶段二.内部攻击者发现外部算力的加入,尚未发现外部算力的攻击行为.内部攻击者将外部算力判定为诚实矿工并重新预测相对收益.

图 7(a)与图 7(b)展示了外部攻击者算力取值 $\frac{1-\gamma_2}{2-\gamma_2}\times\frac{1+\rho}{2}\times(\alpha+\beta_1)$且 β'的取值为 0.25 与 0.45 时,内部攻击者相对收益的预测值与实际值.



(a) $\beta'=0.25$



(b) $\beta'=0.45$

图 7 内部攻击者相对收益的理论值与预测值

图 7 中,内部攻击者两阶段的相对收益预测值都与实际值不相符,这使内部攻击者意识到外来算力并非诚实矿工.表 4 和表 5 更直观地展示了内部攻击者陷入“矿工困境”并错误地预估了相对收益.内部攻击者会在阶段一高估自己的相对收益,误差

最大可达 67%；在 $\beta'=0.25$ 且 γ_1 较高时，内部攻击者在阶段二高估自己的相对收益.

表 4　$\beta'=0.25$ 时外部攻击者的相对收益

γ_1	阶段一理论值	阶段二理论值	实际值
0	0.196(62.65%)	0.104(−13.67%)	0.120
0.2	0.216(62.05%)	0.123(−7.63%)	0.133
0.4	0.239(65.58%)	0.143(−0.46%)	0.144
0.6	0.262(66.14%)	0.165(4.41%)	0.158
0.8	0.282(67.16%)	0.184(9.12%)	0.168
1.0	0.304(67.46%)	0.205(12.75%)	0.181

表 5　$\beta'=0.45$ 时外部攻击者的相对收益

γ_1	阶段一理论值	阶段二理论值	实际值
0	0.657(54.85%)	0.350(−17.41%)	0.424
0.2	0.653(52.68%)	0.368(−14.88%)	0.433
0.4	0.679(52.91%)	0.390(−12.188%)	0.444
0.6	0.685(49.84%)	0.411(−10.14%)	0.457
0.8	0.698(50.93%)	0.430(−7.01%)	0.462
1.0	0.699(46.87%)	0.449(−5.66%)	0.476

5.3　出现“矿工困境”的原因

在两阶段挖矿攻击模型的多攻击者系统中，无论是内部攻击者还是外部攻击者，都对自身的挖矿相对收益产生了错误的预期并陷入“矿工困境”. 在多攻击者系统中，攻击者预期之外的竞争、攻击者拍卖式发布区块与攻击者高估自身影响力是造成攻击者陷入“矿工困境”并对自己的相对收益产生错误预期的主要原因.

攻击者预期之外的竞争. 通常来说，攻击者通过 Match 行为在区块链网络中发布分叉链并产生竞争，在这种情况下区块链网络中产生的竞争为攻击者预期之内的竞争. 但是，在多攻击者系统中，攻击者的分叉链不仅需要与诚实矿工的主链竞争，还需要与另一个攻击者竞争. 这导致了攻击者 Override 行为同样可能导致竞争的出现，该结果与攻击者通过 Override 行为确保自己发布的分叉链成为区块链网络新主链的目的相悖. 因此，本文将攻击者 Override 行为在区块链网络中产生的竞争定义为攻击者预期之外的竞争.

图 8 展示了区块链网络中的两类竞争：攻击者预期之内的竞争与攻击者预期之外的竞争. 在类型一所示的攻击者预期之内的竞争中，某一攻击者(内部攻击者或外部攻击者)持有一个未发布的区块. 在诚实矿工挖到新区块并发布后，攻击者通过 Match 行为发布区块制造分叉链. 此时，在区块链网络中出现了在攻击者预期之内的竞争. 在类型二所示的攻击者预期之外的竞争中，内部攻击者与外部攻击者各自持有两个未发布的区块. 在诚实矿工挖到新区块并发布后，两个攻击者同时通过 Override 行为，试图使自己发布的分叉链成为新的主链. 区块链网络中因此产生了攻击者预期之外的竞争.

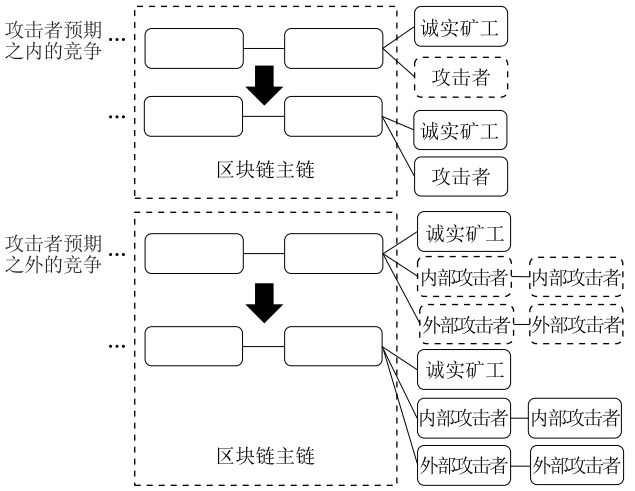


图 8　两类竞争(虚线代表攻击者未发布的区块)

攻击者拍卖式发布区块. 通常来说，攻击者的 Override 行为能够确保攻击者发布的分叉链成为被全区块链网络接受的新主链. 但是，在多攻击者系统中，即使并未出现攻击者预期之外的竞争，攻击者也无法确保自己发布的分叉链成为新的主链. 两个攻击者可能以拍卖的形式，交替发布出比当前主链长一个区块的分叉链. 本文将两个攻击者交替发布更长的分叉链行为定义为攻击者拍卖式发布区块.

图 9 是攻击者拍卖式发布区块的简单场景. 外部攻击者拥有两个尚未发布的区块，内部攻击者拥有三个尚未发布的区块. 当诚实矿工发布新区块后，外部攻击者将首先执行 Override 行为，而内部攻击者由于仅收到了诚实矿工发布区块的信息，执行 Hold 行为. 当内部攻击者最终收到了外部攻击者发

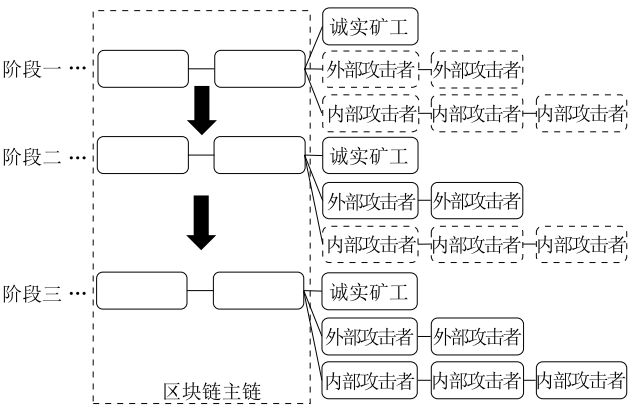


图 9　攻击者拍卖式发布区块(虚线代表攻击者未发布的区块)

布的分叉链后，内部攻击者通过 Override 行为将三个区块全部发布. 在该场景中，外部攻击者执行了 Override 行为，但是由于攻击者拍卖式发布区块，外部攻击者发布的分叉链并未成为主链.

攻击者高估自身影响力. 在多攻击者系统中，攻击者往往高估自身对诚实矿工的影响力 γ_1 与 γ_2. 当区块链网络中的诚实矿工支持的主链与分叉链间存在竞争时，占比 γ_h 的诚实矿工由于网络传播时延的原因，先接收到了攻击者发布的分叉链.

两个攻击者预测自己的相对收益的仿真中，由于内部攻击者与外部攻击者在起初都将对方视作诚实矿工，因此 $\gamma_1=\gamma_h$，$\gamma_2=\gamma_h$.

而在仿真两个攻击者相对收益的实际值时，为了更符合现实情况，本文通过两个步骤计算参数 γ_1 与 γ_2：

(1) 判定竞争类型. 当攻击者参与的竞争为攻击者预期外的竞争时，$\gamma_h=1$；当攻击者参与的竞争为攻击者预期内的竞争时，γ_h 需要在仿真中设定.

(2) 分配. 将参与竞争的所有攻击者算力表示为 $\sum_k \beta_k$. 若内部攻击者参与竞争，$\gamma_1=\gamma_h\times\frac{\beta_1}{\sum_k \beta_k}$，$\gamma_2=\gamma_h-\gamma_1$；若内部攻击者未参与竞争，$\gamma_1=0$，$\gamma_2=\gamma_h$.

6 多攻击者系统中的鲶鱼效应

由于外部攻击者的出现，内部攻击者发起挖矿攻击的收益大幅下降并陷入“矿工困境”. 因此内部攻击者需要寻找一个更优的挖矿攻击策略作为针对外部攻击者的对抗措施以破解“矿工困境”.

一个挖矿攻击策略的最终目的是：防止自身算力被其余攻击者浪费；更好地浪费诚实矿工与其余攻击者的算力. 避免意料之外的竞争与避免攻击者拍卖式发布区块可以有效地实现该目的.

6.1 部分主动发布策略

本文提出了一个全新的挖矿攻击策略：部分主动发布策略(Partial Initiative Release，PIR). 部分主动发布策略是一系列策略的策略集，可以被表示为 $\{PIR_1, PIR_2, PIR_3, \cdots, PIR_n, \cdots\}$. 在策略 PIR_n 中，攻击者状态(攻击者未发布的私有链领先主链的长度)的最大值被限定为 n，当攻击者状态超过 n 时，攻击者将主动发布区块. 图 10 展示了部分主动发布策略集中 PIR_3 的状态机. 当采用 PIR_3 策略的攻击者的状态尚未达到 3 时，攻击者的表现与自私挖矿攻击者相同. 在采用 PIR_3 策略的攻击者状态达到 3 后，若下一个区块仍然由该攻击者发现，那么攻击者将主动发布一个区块；若下一个区块由诚实矿工发现，那么攻击者将发布所有的未发布区块.

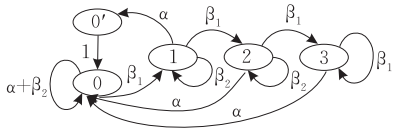


图 10 策略 PIR_3 的状态机

内部攻击者可以通过 PIR_3 策略降低攻击者预料之外的竞争出现的概率与攻击者拍卖式发布区块行为出现的概率.

6.2 仿真量化分析

本文通过仿真揭示在采用自私挖矿策略的外部攻击者出现后，内部攻击者该通过何种策略对抗外部攻击者. 内部攻击者的可选策略有：自私挖矿策略、部分主动发布策略 PIR_3 与诚实挖矿.

在仿真中，本文将内部攻击者算力 β_1 设置为变量，分别探讨外部攻击者算力处于上界 $\beta_2=\frac{1-\gamma_2}{2-\gamma_2}(\alpha+\beta_1)$ 与外部攻击者算力处于下界 $\beta_2=\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)$ 时，内部攻击者采取不同攻击策略时的相对收益. 此外，本文也将讨论不同的 γ_h 取值带来的影响，γ_h 值取自集合 $\{0, 0.25, 0.5, 0, 75\}$.

外部攻击者算力处于下界. 当外部攻击者算力处于下界 $\beta_2=\frac{1-\gamma_2}{2-\gamma_2}\rho(\alpha+\beta_1)$ 时，由于 $\alpha+\beta_1+\beta_2=1$，且 $\rho=\alpha'+\frac{\beta'(\beta'-2\beta'^2)}{(2\beta'^3-4\beta'^2+1)}+\frac{\alpha'\beta'(\beta'-2\beta'^2)}{2\beta'^3-4\beta'^2+1}$，通过 β_1 的值可确定 α 与 β_2 的值. 受模型假设“内部攻击者算力小于诚实矿工”即 $\beta_1<\alpha$ 的约束，仿真中内部攻击者算力小于 1/3. 图 11 展示了内部攻击者采取不同策略时，内部攻击者相对收益与内部攻击者算力 β_1 的关系. 此外图中还将内部攻击者算力作为基线，当内部攻击者相对收益高于内部攻击者算力时，内部攻击者将获得额外攻击收益. 仿真结果表明，当 $\gamma_h=0$ 时，若内部攻击者算力占比大于 0.27，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.27, 0.31]内，部分主动发布策略优于自私挖矿与诚实挖矿. 当 $\gamma_h=0.25$ 时，若内部攻击者算力占比大于0.24，采用部

分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.24,0.28]内，部分主动发布策略优于自私挖矿与诚实挖矿. 当 $\gamma_h=0.5$ 时，若内部攻击者算力占比大于 0.20，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.21,0.26]内，部分主动发布策略优于自私挖矿与诚实挖矿. 当 $\gamma_h=0.75$ 时，若内部攻击者算力占比大于 0.16，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.17,0.23]内，部分主动发布策略优于自私挖矿与诚实挖矿.

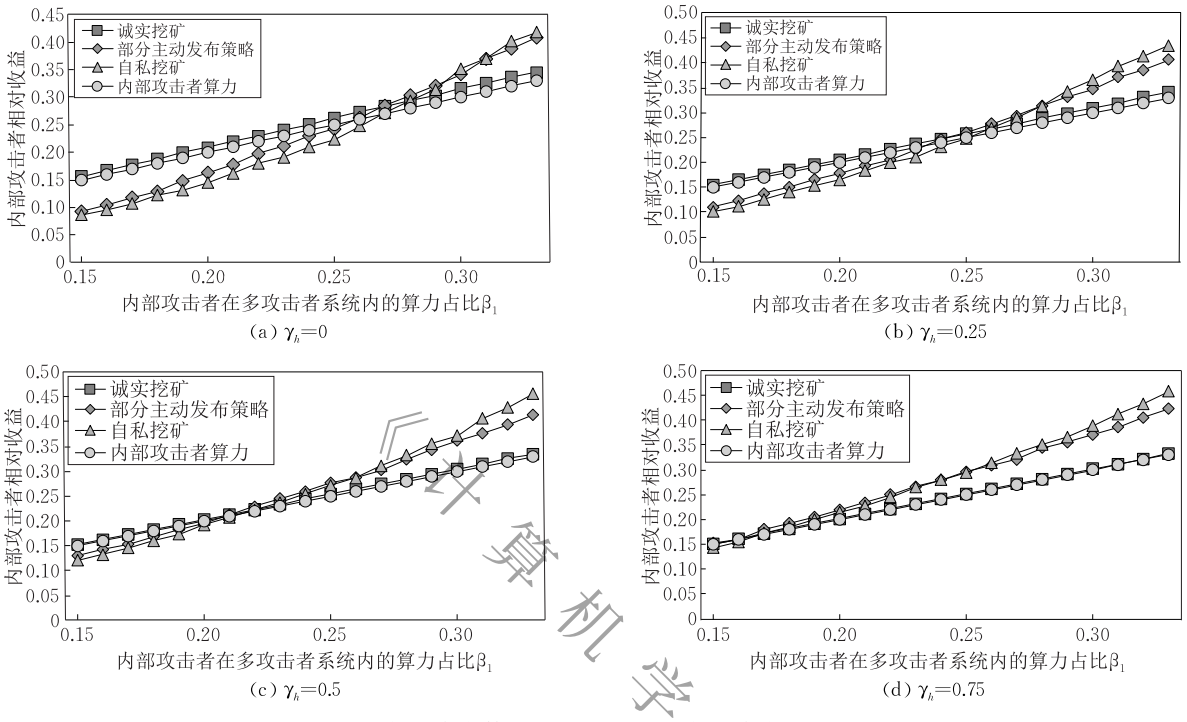


图 11　外部攻击者算力处于下界时，内部攻击者的相对收益

由仿真结果可以看出，在外部攻击者算力较低(处于下界)时，对于内部攻击者来说，部分主动发布策略的优势区间随参数 γ_h 值变化而变化. 相较于自私挖矿策略与诚实挖矿，部分主动发布策略在优势区间内对内部攻击者相对收益的提升有限.

外部攻击者算力处于上界. 当外部攻击者算力处于上界 $\beta_2=\dfrac{1-\gamma_2}{2-\gamma_2}(\alpha+\beta_1)$ 时，由于 $\alpha+\beta_1+\beta_2=1$，通过 β_1 的值可确定 α 与 β_2 的值. 为确保内部攻击者算力小于诚实矿工，在仿真中内部攻击者算力小于 1/3. 图 12 展示了内部攻击者采取不同策略时，内部攻击者相对收益与内部攻击者算力 β_1 的关系. 仿真中，本文将内部攻击者算力作为基线，当内部攻击者相对收益高于内部攻击者算力时，内部攻击者将获得额外的攻击收益. 仿真结果表明，当 $\gamma_h=0$ 时，若内部攻击者算力占比大于 0.25，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.24,0.33]内，部分主动发布策略优于自私挖矿与诚实挖矿. 当 $\gamma_h=0.25$ 时，若内部攻击者算力占比大于 0.24，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.23,0.31]内，部分主动发布策略优于自私挖矿与诚实挖矿. 当 $\gamma_h=0.5$ 时，若内部攻击者算力占比大于 0.21，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.22,0.27]内，部分主动发布策略优于自私挖矿与诚实挖矿. 当 $\gamma_h=0.75$ 时，若内部攻击者算力占比大于 0.15，采用部分主动发布策略将为内部攻击者带来额外收益；若内部攻击者算力占比处于区间[0.15,0.24]内，部分主动发布策略优于自私挖矿与诚实挖矿.

由仿真结果可以看出，在外部攻击者算力较高(处于上界)时，对于内部攻击者来说，部分主动发布策略的优势区间随参数 γ_h 值变化而变化. 当 γ_h 较低($\gamma_h=0$ 或 $\gamma_h=0.25$)时，相较于自私挖矿策略与诚实挖矿，部分主动发布策略在其优势区间内对内部攻击者的相对收益有显著的提升. 当 γ_h 较高($\gamma_h=0.5$ 或 $\gamma_h=0.75$)时，相较于自私挖矿策略与诚实挖

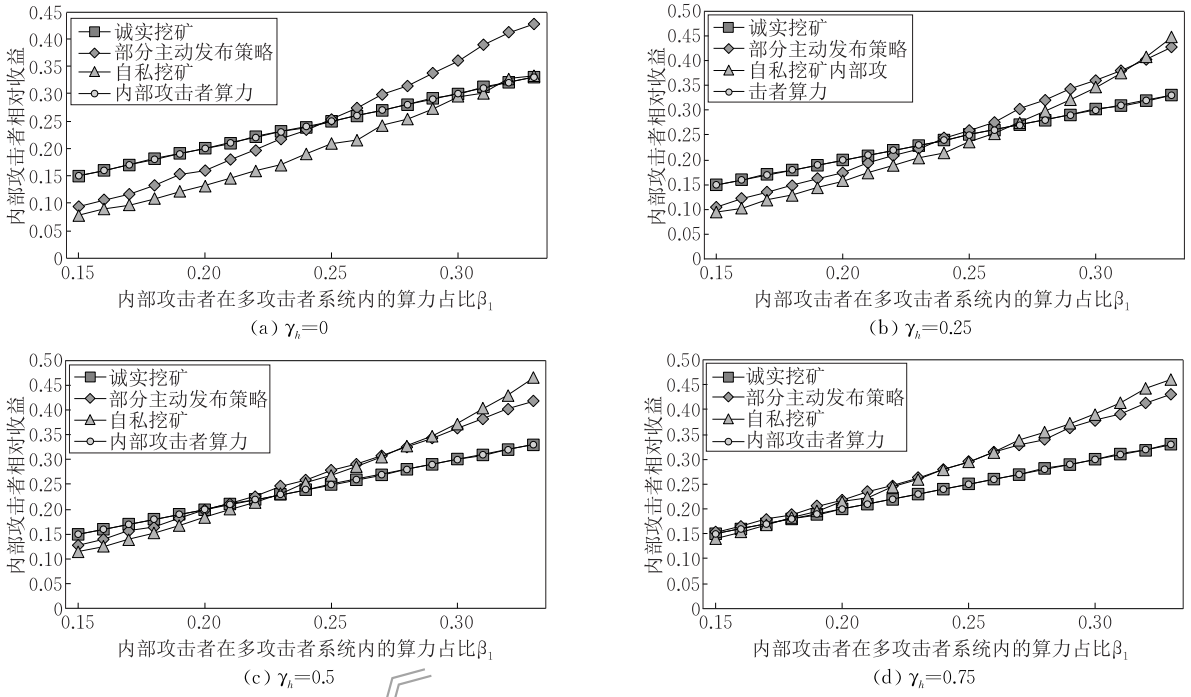


图 12 外部攻击者算力处于上界时,内部攻击者的相对收益

矿,部分主动发布策略在其优势区间内对内部攻击者的相对收益提升有限.

仿真结果表明,在多攻击者系统中,内部攻击者可以依据自身算力 β_1 与外部攻击者算力 β_2 选择攻击策略以对抗外部攻击者.当外部攻击者算力较高时,内部攻击者采用部分主动发布策略可以显著提升自身的相对收益;而当外部攻击者算力较低时,部分主动发布策略对内部攻击者相对收益的提升有限.

7 结 论

本文提出了全新的两阶段挖矿攻击模型并指出当挖矿攻击模型中存在多个攻击者时,攻击者将面临"矿工困境".两阶段挖矿攻击模型由传统自私挖矿系统与多攻击者系统组成.本文解释了挖矿攻击模型出现两个攻击者的原因与攻击者们陷入"矿工困境"的原因.本文还揭示了多攻击者系统中存在着鲶鱼效应:在多攻击者系统中,由于外部攻击者的出现带来了预期之外的竞争问题与攻击者拍卖式发布区块问题,内部攻击者需要优化攻击策略.本文提出了名为部分主动发布策略的半诚实挖矿攻击策略,在部分场景下,攻击者可以通过部分主动发布策略获得更高的攻击收益,从而打破"矿工困境".此外,尽管本文从内部攻击者的视角提出了部分主动发布策略,实际上外部攻击者同样可以使用该策略提升自身的相对收益.

参 考 文 献

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Bitcoin.—URL: https://bitcoin.org/bitcoin.pdf, 2008

[2] Gervais A, Karame G O, Wüst K, et al. On the security and performance of proof of work blockchains//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna, Austria, 2016: 3-16

[3] Eyal I. The miner's dilemma//Proceedings of the 2015 IEEE Symposium on Security and Privacy. San Jose, USA, 2015: 89-103

[4] Eyal I, Sirer E G. Majority is not enough: Bitcoin mining is vulnerable. Communications of the ACM, 2018, 61(7): 95-102

[5] Nayak K, Kumar S, Miller A, et al. Stubborn mining: Generalizing selfish mining and combining with an eclipse attack//Proceedings of the 2016 IEEE European Symposium on Security and Privacy. Saarbrücken, Germany, 2016: 305-320

[6] Sapirshtein A, Sompolinsky Y, Zohar A. Optimal selfish mining strategies in Bitcoin//Proceedings of the International Conference on Financial Cryptography and Data Security. Christ Church, Barbados, 2016: 515-532

[7] Tang Chang-Bing, Yang Zhen, Zheng Zhong-Long, et al. An analysis and optimization on miner's dilemma in PoW consensus protocol. Acta Automatica Sinica, 2017, 43(9):

1520-1531(in Chinese)
(唐长兵，杨珍，郑忠龙等. PoW 共识算法中的博弈困境分析与优化. 自动化学报，2017，43(9)：1520-1531)

[8] Kwon Y, Kim D, Son Y, et al. Be selfish and avoid dilemmas: Fork after withholding (FAW) attacks on Bitcoin//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. Dallas, USA, 2017: 195-209

[9] Gao S, Li Z, Peng Z, Xiao B. Power adjusting and bribery racing: Novel mining attacks in the Bitcoin system//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. London, UK. 2019: 833-850

[10] Dong X, Wu F, Faree A, et al. Selfholding: A combined attack model using selfish mining with block withholding attack. Computers & Security, 2019, 87: 101584

[11] Liu H, Ruan N, Du R, et al. On the strategy and behavior of Bitcoin mining with N-attackers//Proceedings of the 2018 on Asia Conference on Computer and Communications Security. Incheon, South Korea, 2018: 357-368

[12] Marmolejo-Cossío J, Brigham E, Sela B, Katz, J. Competing (Semi-) selfish miners in Bitcoin//Proceedings of the 1st ACM Conference on Advances in Financial Technologies. New York, USA, 2019: 89-109

[13] Bai Q, Zhou X, Wang X, et al. A deep dive into blockchain selfish mining//Proceedings of the 2019 IEEE International Conference on Communications. Shanghai, China, 2019: 1-6

[14] Merkle R C. A digital signature based on a conventional encryption function//Proceedings of the Conference on the Theory and Application of Cryptographic Techniques. Berlin, Germany, 1987: 369-378

[15] Laszka A, Johnson B, Grossklags J. When Bitcoin mining pools run dry//Proceedings of the International Conference on Financial Cryptoy and Data Security. San Juan, Puerto Rico, 2015: 63-77

[16] Decker C, Wattenhofer R. Information propagation in the Bitcoin network//Proceedings of the 2013 IEEE P2P. Trento, Italy, 2013: 1-10

[17] Bag S, Sakurai K. Yet another note on block withholding attack on Bitcoin mining pools//Proceedings of the International Conference on Information Security. Honolulu, USA, 2016: 167-180

[18] Bag S, Ruj S, Sakurai K. Bitcoin block withholding attack: Analysis and mitigation. IEEE Transactions on Information Forensics and Security, 2016, 12(8): 1967-1978

[19] Heilman E, Kendler A, Zohar A, et al. Eclipse attacks on Bitcoin's peer-to-peer network//Proceedings of the 24th USENIX Security Symposium. Washington, USA, 2015: 129-144

[20] Göbel J, Keeler H P, Krzesinski A E, et al. Bitcoin blockchain dynamics: The selfish-mine strategy in the presence of propagation delay. Performance Evaluation, 2016, 104: 23-41

[21] Bonneau J, Miller A, Clark J, et al. SoK: Research perspectives and challenges for Bitcoin and cryptocurrencies//Proceedings of the 2015 IEEE Symposium on Security and Privacy. San Jose, USA, 2015: 104-121

[22] Liao K, Katz J. Incentivizing blockchain forks via whale transactions//Proceedings of the International Conference on Financial Cryptography and Data Security. Sliema, Malta, 2017: 264-279

RUAN Na, Ph. D., associate professor. Her research interests include security & privacy, blockchain and big data.

LIU Han-Qing, M. S. His research interests include security of blockchain.

SI Xue-Ming, professor. His research interests include cryptography, data science, computer architecture, network security and blockchain.

Background

Most digital cryptocurrencies are based on the blockchain technology. The consensus mechanism of a blockchain guarantees that the records, especially transaction records stored in the blockchain are irreversible and consistent to all users. Proof of Work, Proof of Stake and Practical Byzantine Fault Tolerance are some consensus mechanism which are applied by blockchain. Among these consensus mechanisms, Proof of Work is the most widely used one. In the past decade, the security of a Proof of Work powered blockchain has been a hot topic. Researchers have proposed many mining attack strategies such as selfish mining attack and block withholding attack. These mining attack strategies are proved to be feasible not only by the researchers, but also by the attackers in the real world. For example, in May 2018, a Proof of Work powered cryptocurrency named Monacoin was attack by an attacker with the famous mining attack strategy

selfish mining.

Mining attacks can be prevented by honest miners via designing appropriate protocols. But an interesting fact is that mining attacks can also be prevented by attackers themselves. Some attackers have to turn into honest miners due to the competitions between attackers. Some researches reveal that attackers might face prisoner's dilemma when they are conducting block withholding attack. To some extent, the game between attackers decreases all attacker's reward. But the competition between attackers with strategy selfish mining is not systematically studied by the researchers. Our work named 'On the strategy and behavior of Bitcoin mining with N-attackers' presented the first selfish mining model with two attackers in 2018. On the basis of our previous work, in this work, we present a two-phase mining attack model and reveals the Catfish effect between two selfish miners. Our work shows that in the mining attack model with two attackers, being semi-honest is an option for the attacker. Thus, when the number of attacker increases, attackers have the trend to be honest or semi-honest.

This work is supported by the National Natural Science Foundation of China under Grant No. 61702330.