$$R=\begin{cases}\max\sum\limits_{V_i\in U_{attack}} child(V_i), & \sum\limits_{i=1}^{n}V_i\geqslant 1\\ \max\sum\limits_{V_i\in U_{all}} child(V_i), & \sum\limits_{i=1}^{n}V_i=0\end{cases} \tag{2}$$

R 表示攻击者在当前区块链中选择连接的区块,V_i 是区块链中的每个区块,函数 $child$ 用来判断节点 V_i 是否有子节点.当区块中存在攻击区块时,攻击者会将新的区块连接到攻击区块后连接区块最长的链.如果不存在攻击区块,则选择最长的分支.

图 1 是从攻击者的角度看到的区块链的状态图,实心节点 c 表示该区块含有虚假的交易信息,其余节点是验证正确的区块.在此状态下,攻击者会将新产生的区块连接到 c 节点后,这样会使 c 节点的交易信息更不会被更改,加大攻击成功的概率.c 节点后面连接的区块越多,内容就越不容易被更改,当一个含有虚假交易信息的区块到达安全状态时,即认为攻击成功.无论攻击的类型是什么,攻击者的目的都类似.本文使用 51%攻击作为唯一的攻击类型,当攻击类型改变时,也可以采取类似的方法,仅改变普通状态到攻击成功状态概念计算方法即可.

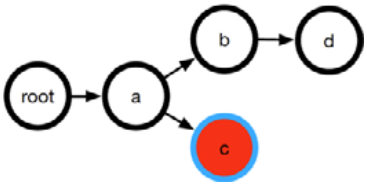

Fig.1 An attackers' strategy

图 1 攻击者策略

2.3 诚实矿工策略定义

对于诚实矿工来说,每个区块的类型是未知的[15].在区块链中,系统只认可最长的链的区块中的交易信息.从概率层面分析,诚实的矿工可以将新产生的节点连接到任意区块下,但是会尽可能地选择最长链的叶子节点.当有多个叶子节点所属链的长度相同时,攻击者将以相同的概率连接到其中的某个节点上.每当节点的深度递减一层,该节点被选择的概率就会降低一半.公式(3)中所有节点被选中的概率之和等于 1,用公式(4)来计算节点 p_{ij} 被选中的概率与它所处的树的层次的关系.

$$\sum_{i=1}^{n}\sum_{j=1}^{m}(1/2)^{(L-i)}p=1 \tag{3}$$

$$p_{ij}=(1/2)^{(L-i)}p \tag{4}$$

L 是整个区块链的长度;P 为最长链的叶子节点被选中的概率,会随着区块链的状态不同而改变.在实际运行环境中,诚实矿工选择叶子节点之前的节点的概率将会更低.为了模拟方便,将递减的概率设置为 0.5[8].诚实攻击者选择的策略比攻击者更复杂,因为它们无法获知每个区块的状态.越靠近根节点,被选择的概率越低.在图 1 的结构下,节点 d 被选中的概率最大,该概率只与节点的深度有关.

3 实验方法

本节将介绍区块链系统模拟算法、攻击策略,并用数学的方法证明区块链中的状态数目是有限的.区块链是一个十分庞大和复杂的去中心化的系统,它用数学的方法解决了双方之间的信任问题.但其中仍然存在很多问题,例如双重支付、51%攻击、日蚀攻击等.本文使用模拟的方法获得整个过程的状态,并对每个状态进行分析,从而评估整个系统的安全性.

3.1 挖矿过程模拟算法

本文采用树型结构来代表区块链.区块链中的节点有两种类型:攻击节点或诚实节点.设置一个概率代表攻击力度,即攻击者拥有的计算资源比例,攻击力度与下一个新区块的类型有关.根据不同的区块的类型,将采用

不同的连接策略来选择连接的节点.算法 1 是挖矿过程模拟的算法.

算法 1. 获得区块链模拟过程中的全部状态.

输入:攻击力度 *P*.

输出:区块链中的所有状态 *S*.

1. 创建一个诚实节点作为区块链的初始化过程
2. 循环
3. 根据攻击力度创建一个新的区块
4. 如果新的区块为诚实节点
5. 使用算法 2 //诚实矿工策略选择下一个连接节点
6. 否则
7. 选择所有攻击节点所在最长链的叶子节点进行连接,若没有攻击节点,则选择最长链的叶子节点.
8. 将新的节点连接到选择的节点上
9. 如果该状态与已出现的状态都不相同
10. 将该状态保存在集合 *S* 中
11. 如果一个结构变成了安全状态或者攻击状态
12. 重新用一个诚实节点初始化区块链
13. 当区块链中的状态数目收敛时,停止循环
14. 返回模拟过程中所有的状态集合 *S*

算法 1 忽略了工作量证明和时间戳服务等,因为这些复杂的过程和我们的研究目标没有联系.模拟过程中只要有 1 个区块达到稳定状态,则该状态之前的所有节点都可以被剪掉,因为它们的内容都无法被更改,从而降低问题的复杂程度.在每次循环中,我们都会将产生的状态与已有的区块状态进行比较,如果两个状态中的树的分支相同,包括节点数目、分支深度等,则这两个状态相似.因为分支深度有限,且根据协议特征,区块链中所有的状态数目与攻击力度、攻击方式无关.

3.2 证明:区块链中的状态是有限的

根据中本聪的白皮书中所提到的,攻击者追上 *z* 个区块的概率为

$$\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}\begin{cases}(q/p)^{(z-k)}, & k\leqslant z\\ 1, & k>z\end{cases}.$$

使用该公式,攻击者在拥有 30%的攻击力度,且攻击区块比诚实区块落后 5 个区块时,攻击者能够将攻击区块所在的链变成最长链的概率小于 18%.实际上,10%的攻击算力写的时间为 100GHs,因此,如果一个区块后面连接 6 个区块,则该区块是足够稳定的.以上证明了区块的深度是有限的.

根据诚实矿工的策略,当一个根节点只有 1 个叶子节点时,连接到根节点的概率为 2/3.下一个区块仍然连接到根节点的概率降低到 2/5.以此类推,一个节点有 5 个分支的概率小于 0.017.因此,区块中的所有节点的分支将小于等于 5,即分支的数目也是有限的.

总的来说,模拟过程中的状态小于等于 $\frac{a_n q-a_1}{q-1}=\frac{5^7-5}{5-1}=19530.$

出现其他情况的概率极小,可以不予考虑.

3.3 攻击者和诚实矿工的挖矿算法

挖矿是区块链中十分重要的过程,每个人都可以参与和维护它的运行.根据协议,诚实矿工倾向于接受最新的交易信息,并将最长链上的交易作为可信的记录.如果在同一时间段中出现两个区块,则区块链在短时间内会出现不一致的现象.新产生的区块可以任意选择想要连接的区块,与攻击者行为类似,因此无法区分攻击者和诚实者.算法 2 将诚实矿工的行为进行简化和模拟,实际上诚实矿工的行为更加复杂.该算法描述绝大多数情况下,

诚实矿工的行为.即链的深度越深,该链被选择的概率越大;随着深度的减少,被选择的概率不断降低.

算法 2. 选择诚实区块连接的节点.

输入:区块链中某个状态的树状结构的根节点 *r*.

输出:一个诚实区块将会连接的节点.

1. 计算树的深度
2. **for** 树的层数 *i*=0 到 *n*
3. **for** 第 *i* 层的每个节点 *j*=0 to *m*
4. 将每个节点权重加合
5. 根据计算得到的概率 *p*,选择一个连接的节点 *S*
6. 返回节点 *S*

对于攻击者,他们清楚整个系统中每个区块的状态,因此,他们会尽可能地将新产生的区块连接到自己交易所在的链上.如果有多个攻击区块在不同链上,则选择攻击区块后面节点最多的叶子节点进行连接.如果没有攻击区块,则选择最长链的叶子节点进行连接.攻击者可以使用很多攻击方法甚至是物理攻击来修改交易信息.我们很难去将所有的攻击进行模拟,但是众多攻击的目标相同,因此仅仅需要用一个简单的模型来模拟和分析这个过程.在第 3.1 节中证明了区块链的状态是有限的,则分析每个状态安全或者被攻击成功的概率是可行的.

4 实验和结果分析

本节使用不同的方法来展示实验的结果,分析模拟循环次数、状态数目、攻击数目之间的关系.同时,用一些真实的例子来展示方法的可行性.

4.1 区块链的状态数目与攻击力度的关系

实验在一台配置有 40G 内存、1.8GHz cpu 的 linux 服务器上完成.将攻击力度设置为 10%~60%改变,来统计状态数目.

第 3.1 节中证明了区块链的状态数目小于 19 530,超过的概率极小.因此,本实验设置的循环次数为 25 000.循环次数越多,获得的状态数目会越多.如果想获得全部状态,需要运行很长的时间且可行度低.本实验只需获得区块链绝大多数状态就足够.

图 2 是循环次数为 25 000 时,不同攻击力度下,区块链状态的数目.图 2 的最大值为 19 325,与我们在第 3.1 节中证明的状态数目小于 25 000 吻合.从图中可以发现:不同攻击力度下,状态数目都接近 19 300.结果表明,系统状态数目与攻击力度无关.在该模拟过程中,状态数目值在 19 300 附近,因此,后面的实验将 19 300 作为状态总数.

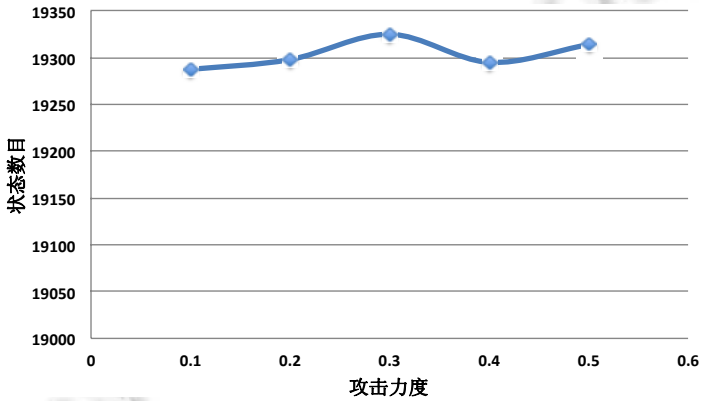

Fig.2 Variation of state number in different attacking power

图 2 不同攻击力度下,状态数目的变化

通过进一步的实验,我们研究分析了循环次数、攻击数目和状态数目这三者之间的关系.实验过程将循环次数设置为 1 000~27 000,然后统计状态数目、攻击数目和循环次数,并用折线图来展示实验的结果.攻击状态是指当一个包含虚假交易信息的区块后面连接了 6 个及以上的区块后,则该区块为攻击区块,该区块链处于攻击成功状态.图 3 展示了状态数目、攻击数目和循环次数之间的关系.从图中可以发现:随着循环次数的增加,状态数目和攻击数目也逐渐增加,状态数目增加的速率比攻击数目增加的速率快.

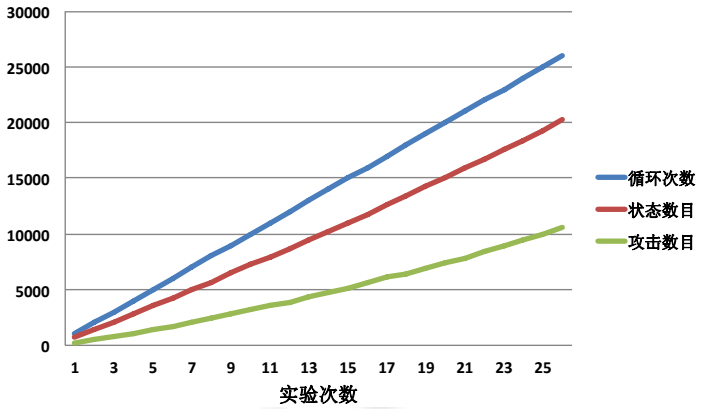

Fig.3 Relationship between state number, attacking number and cycle times

图 3 状态数目、攻击数目和循环次数的关系

图 4 展示了不同循环次数下,状态数目和攻击数目的增长率之间的关系.从图 4 中发现:当循环次数较小时,两者的增加率较高;随着循环次数的不断增大,增长率不断下降;当循环次数达到 17 000 以上后,增长率基本不改变,维持在 0.05.从图中可以推测出:当循环次数继续增大时,增加率会缓慢下降,直到 0 为止,攻击数目和状态数目都会收敛于一个稳定值.

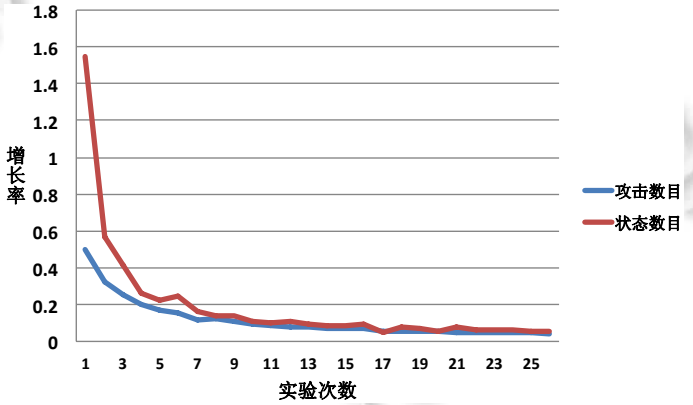

Fig.4 Growth rate of state number and attacking number

图 4 攻击数目和状态数目增长率

4.2 区块链的攻击数目与攻击力度的关系

本文的主要目的是分析攻击的特征,评估每个状态区块链的安全性.因此,我们设置实验条件为循环次数为 20 000~26 000,攻击力度为 0.1~0.6,分析攻击数目和攻击力度之间的关系.

图 5 中横坐标表示攻击的力度,纵坐标为攻击的数目.随着攻击力度的增加,攻击数目也逐渐增加.即攻击者掌控的计算资源越多,发生攻击的数目就会越大.不同的直线表示不同的循环次数,我们使用最小二乘法计算攻

击数目变化趋势.从图中可以发现:随着循环次数的增加,攻击数目的增加速率降低.即循环次数越大,攻击数目变化越小.

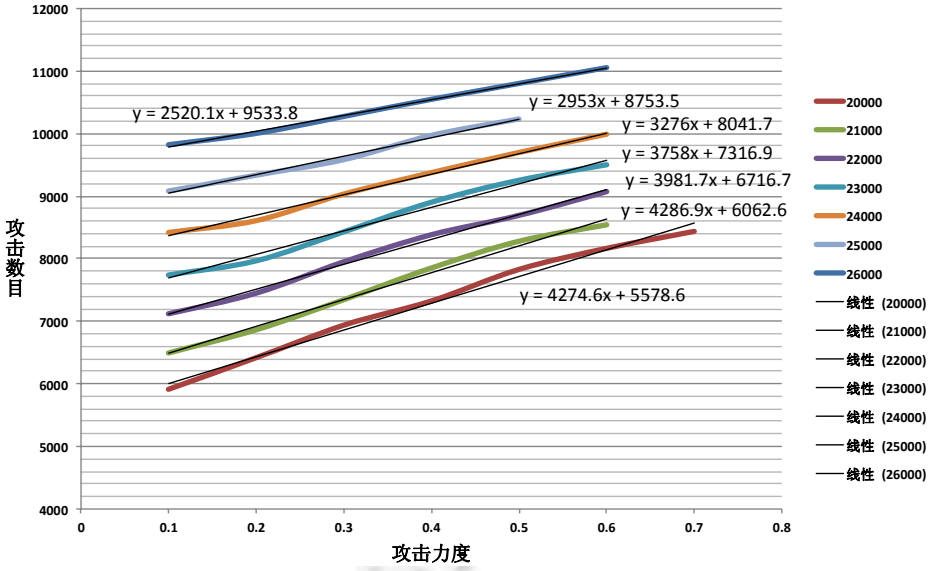

Fig.5 Relationship between attacking number and cycle times

图 5 攻击数目和循环次数的关系

从图 6 中也可以发现:当攻击力度固定时,攻击数目的最大和最小值之间的差值随着循环次数的增加变得越来越小.总的来说,图 5、图 6 的结果表明:攻击数目与攻击力度、循环次数正相关,且变化速率不断减小,并将收敛于某个值.第 4.1 节中证明了区块链总的状态数目是恒定的,本节用实验证明了攻击数目也是收敛的.这些特征说明,分析每个状态到攻击状态的概率是可行的,因为它们的数量是有限的.同时我们发现,状态数目和攻击力度之间的关系不是简单的线性关系.

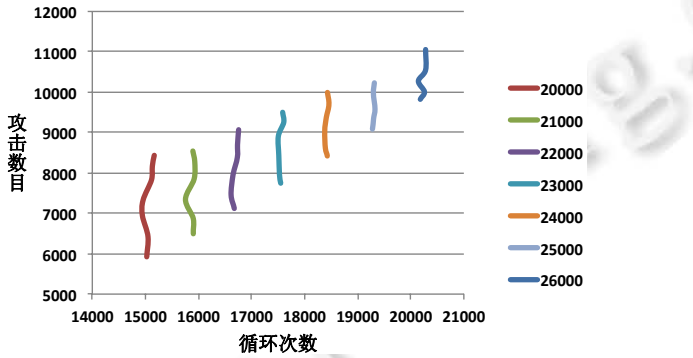

Fig.6 Gap between the maximum and minimum number of attacks

图 6 攻击数目最大值与最小值之间的间距关系

图 7 是循环次数为 24 000、攻击力度为 0.1~0.6 时,攻击力度和状态数目的关系.横坐标为攻击数目,纵坐标是状态数目.

从图 6 的实验结果可以发现:两者的关系与正弦函数类似,即两者呈现周期性的变化.再次改变循环次数为 20 000~25 000,攻击数目和状态数目之间的关系仍然呈现周期性变化.

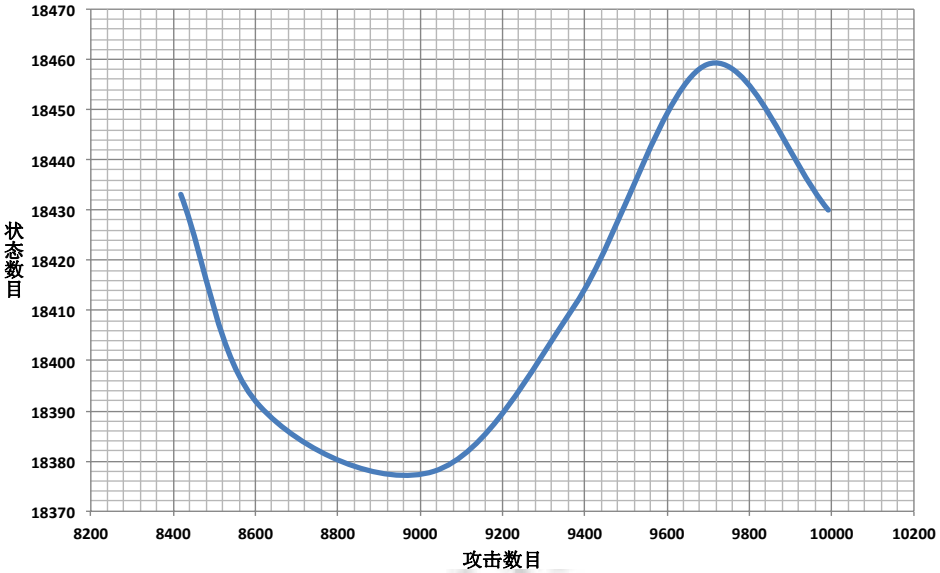

Fig.7 Relationship between state number and attacking number in different attacking power

图 7 不同攻击力度下,状态数目和攻击数目的关系

4.3 每个状态被攻击的概率计算

通过将实验中的循环次数设置为 25 000、攻击力度设置为 0.4 来分析区块链的深度和状态数目之间的关系.表 1 统计了区块链不同层次中状态的数目和攻击成功的数目.实验结果表明:总的攻击数目为 832,深度为 6 的状态数目为 17 356.两者的比例 8732/17356 近似于 0.5,符合我们的预期.因为诚实攻击者的选择策略中概率和区块链的深度之间呈现一种等比关系,且最后一层的状态数目占总的状态数目的一半,攻击数目的变化规律与状态数目的变化类似.

Table 1 State number in different depth of blockchain

表 1 不同深度区块链的状态数目

深度	状态数目	攻击数目
1	1	1
2	15	6
3	113	19
4	871	56
5	6 644	1 120
6	17 356	8 732

根据模拟算法,我们获得了区块链中各个层次中的状态和攻击结构,从而可以分析不同层次的状态到达各个攻击状态的概率,评估系统的安全性.

我们把找到的攻击状态作为分析的目标,分析不同结构到达攻击状态结构的概率.这些概率的总和代表该状态下被攻击成功的可能性大小.当该值大于设定的参数值时,可以向区块链中的用户发送警告,采取一定的措施来降低攻击概率,例如延长确认交易的时间等.图 8 是实验过程中发现的攻击状态的结构特征.图 8(a)中,节点 3 包含了虚假交易,诚实矿工不知道区块的状态,仍然可能将新产生的区块连接上去.攻击者则会尽快地将更多的节点连接到节点 3 所在的链上,攻击者也可以采取日蚀攻击,将找到的区块存放一段时间后一次性释放,使区块的交易信息尽快处于稳定状态.攻击类型很多,从攻击层面分析系统的安全性将会比较复杂,而区块链中的结构是相同且有限的,通过分析每个状态的结构特征,计算被攻击的概率,可以评估系统的安全性.例如,我们有图 8(b)结构的特征,当该结构的高度为 3 时,我们使用去随机化的方法计算该结构到高度为 4 的攻击结构的概率,由表 1 可以发现:深度为 4 时的状态有 871 个,深度为 3 的其他状态有 112 个.可以计算该状态到这 983 个状态

的概率来减少计算状态数目,因为很多状态无法从当前的状态添加一个节点后获得.以此类推,直到计算出到达攻击状态的概率为止,从而获得攻击的可能性大小,评估系统的安全性.

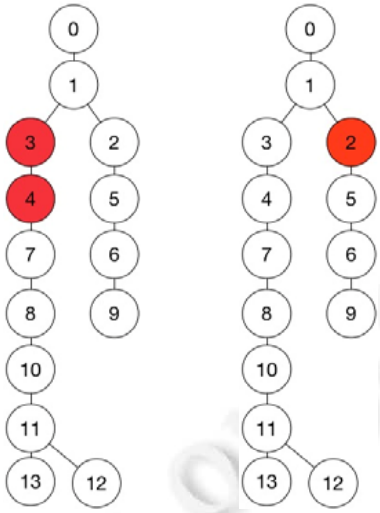

Fig.8 Structure of attacking state

图 8 攻击状态结构

在真实情况下,如果顾客 *A* 创建一个交易,将一定数量的比特币发送给 *B*,从商家 *B* 获得服务;同时,顾客 *A* 又创建一个交易,使用相同的比特币从商家 *C* 获得其他商品.当工具检测到有比特币被花费多次时,可以检测当前状态下结构,对每种情况进行分类讨论,即分第 1 笔交易是虚假的和第 2 笔交易为虚假的两种情况进行分析,计算出两种情况下被攻击的概率.选择概率较低的情况,将另一个情况中的交易进行标记,系统不再承认该交易,同时向用户发送提醒,采取措施来降低攻击概率.

5 相关工作

目前,对区块链安全性的研究大多数是通过定义一些数学模型,然后改变其中的参数来分析结果.Emmanuelle 等人[12]形式化定义了区块链的正确性(validity property)、双重支付状态(double-spending situation)、正确交易(conflict-free transation)和验证过程(local confirmation),构造了区块链模型.通过数学推导,证明了区块链的 4 种性质与矿工有关,分析了与区块链安全相关的因素和提高安全性的方法.但是这些都是理论上的证明和分析,实际运行过程中环境更为复杂,因此需要结合区块链的实际运行状态来分析安全性.

本文与 Emmanuelle 研究的区别主要在于:Emmanuelle 将整个区块链的运行过程当成一个统一的整体,而本文将区块链分成一个个循环,分析从初始状态到攻击状态的过程,从而分析系统的安全性,并用应用程序来分析区块链的安全性.

Rafael 等人[16]用(h_{-1},n,m)表示链状结构,h_{-1} 是指向区块链中前一区块的指针,m 是区块链标识,n 代表工作量证明.并用 p 表示挖坑的难度,构造出一个简化的区块链协议模型.同时,规定诚实矿工在传递消息时,必须延长 L 段时间来模拟网络延迟的情形,从而分析区块链协议在异步网络情形中的安全性.在本文中,攻击者和诚实矿工都及时进行消息传递,且诚实矿工和攻击者的数目是确定的.通过获得区块链的状态来分析区块链的安全,在进一步研究中,可以考虑网络延迟的情况.

Arthur 等人[7]结合区块链中的具体实例,如比特币、莱特币(Litecoin)、Dogecoin 和 Ethereum,根据不同实例的特征构建了不同的模型,并通过改变模型中的参数来分析底层协议区块链的性能.Arthur 用 r_s 阻塞率表示

区块大小、区块时间间隔网络延迟、信息传递机制等影响,用 a 表示攻击者掌握的算力,Cm 表示挖坑的代价,例如硬件、电力和人力等.K 表示一个交给被 k 个用户接受.用 $M\langle S,A,P,R\rangle$表示单个用户的决策问题,S 是状态空间,A 是动作空间,P 表示转移矩阵,R 是奖励矩阵,M 表示马尔可夫决策过程.通过模型参数的改变,分析出将区块大小改成 1MB,间隔时间缩短到 1min,不会严重影响系统的安全性,且系统吞吐量会高于每秒 60 个事务 *tps*. Arthur 等人分析区块大小和时间间隔对区块链安全性的影响,而本文不针对具体的实例进行分析,更关注于区块链本身协议的特征.

与已有的研究方法进行比较,本文将区块链过程划分成循环的过程,构建了模型来模拟区块链的运行过程,找到区块链的状态,通过研究状态的转变来评估系统的安全性.

6 总结以及未来的工作

本文提出了一种树状结构的方法来模拟区块链的运行过程,并分析攻击数目和状态数目的关系,以此来计算区块链中每个状态的安全性.该方法具有通用性,任何的攻击影响都可以使用该方法来评估.本文使用了 51%的算力攻击作为唯一的攻击方式,通过简化区块链中复杂的过程,包括工作量证明、时间戳等,用树状结构来代替区块链的状态,研究不同攻击力度、循环条件下,攻击数目、状态数目之间的关系,并用实验进行验证.

该方法也存在一些局限性.

- 我们采用唯一的一种攻击方式 51%攻击来进行模拟,虽然我们认为攻击策略和状态等无关,但没有实验证明攻击类型和状态数目、攻击数目之间的联系.
- 同时,真实区块链中诚实矿工选择策略会更加复杂,没有恒定的规律,本文仅用递减的模型来模拟诚实矿工的行为.后续可以继续进行改进和研究.

区块链被认为是 21 世纪极具前景的技术,已经被应用在去中心化金融交易平台、智能合约、物理跟踪等领域.比特币平台不断有各种攻击产生,这也是比特币还未在全球使用的原因之一.区块链的安全性亟需一种方法来进行分析和检测,降低被攻击的可能性.未来可以实时地监控区块链系统的结构,并将该检测工具应用于实际的安全性检测中.

References:

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Consulted, 2008.

[2] Grinberg R. Bitcoin: An Innovative Alternative Digital Currency. Social Science Electronic Publishing, 2011.

[3] Nayak K, Kumar S, Miller A, Shi E. Stubborn mining: Generalizing selfish mining and combining with an eclipse attack. In: Proc. of the 2016 IEEE European Symp. on Security and Privacy (EuroS P). 2016. 305–320. [doi: 10.1109/EuroSP.2016.32]

[4] Eyal I, Gencer AE, Sirer EG, Van Renesse R. Bitcoin-NG: A scalable blockchain protocol. In: Proc. of the 13th USENIX Symp. Networked Systems Design and Implementation (NSDI 16). USENIX Association, 2015. 45–59.

[5] Bissias G, Levine BN, Ozisik AP, Andresen G. An analysis of attacks on blockchain consensus. arXiv preprint arXiv:1610.07985, 2016.

[6] Natoli C, Gramoli V. The balance attack against proof-of-work blockchains: The R3 testbed as an example. arXiv preprint arXiv:1612.09426, 2016.

[7] Heilman E, Kendler A, Zohar A, Goldberg S. Eclipse attacks on bitcoin's peer-to-peer network. In: Proc. of the Usenix Conf. on Security Symp. 2015. 129–144.

[8] Garay J, Kiayias A, Leonardos N. The bitcoin backbone protocol: analysis and applications. In: Proc. of the Annual Int'l Conf. on the Theory and Applications of Cryptographic Techniques. Berlin, Heidelberg: Springer-Verlag, 2015. 281–310. [doi: 10.1007/978-3-662-46803-6_10]

[9] Lewenberg Y, Sompolinsky Y, Zohar A. Inclusive block chain protocols. In: Proc. of the FC. 2015. 528–547. [doi: 10.1007/978-3-662-47854-7_33]

[10] Decker C, Wattenhofer R. Information propagation in the bitcoin network. In: Proc. of the IEEE 13th Int'l Conf. on Peer-To-Peer Computing. 2013. 1–10. [doi: 10.1109/P2P.2013.6688704]

[11] Gervais A, Karame GO, Glykantzis V, Ritzdorf H, Capkun S. On the security and performance of proof of work blockchains. In: Proc. of the ACM Sigsac Conf. on Computer and Communications Security. 2016. 3–16. [doi: 10.1145/2976749.2978341]

[12] Yermack D. Corporate governance and blockchains. Review of Finance, 2017,21(1):7–31.

[13] Kaskaloglu K. Near zero bitcoin transaction fees cannot last forever. In: Proc. of the Society of Digital Information and Wireless Communication. 2014. 91–99.

[14] Liehuang Z, Feng G, Meng S, Yandong L, Baokun Z, Hongliang M, Zhen W. Survey on privacy preserving techniques for blockchain technology. Journal of Computer Research and Development, 2017,54(10):2170–2186.

[15] Kumar A, Fischer C, Tople S, Saxena P. A traceability analysis of monero's blockchain. In: Proc. of the European Symp. on Research in Computer Security. 2017. 153–173.

[16] Vasek M, Thornton M, Moore T. Empirical analysis of denial-of-service attacks in the bitcoin ecosystem. In: Proc. of the Financial Cryptography and Data Security. Berlin, Heidelberg: Springer-Verlag, 2014. 57–71. [doi: 10.1007/978-3-662-44774-1_5]

叶聪聪(1993－),女,湖北黄石人,硕士生,主要研究领域为软件工程,安全检测,管理信息系统.

蔡鸿明(1975－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为协同计算,服务计算, CAD/CG,语义数据处理,智慧制造.

李国强(1979－),男,博士,副教授,CCF 专业会员,主要研究领域为形式化方法,理论计算机科学,程序语言理论.

顾永跟(1968－),男,博士,教授,CCF 高级会员,主要研究领域为计算经济学,物联网.