GHOST 阅读笔记

文章概括：

这篇文章主要内容可以分为三块：

1. 为什么比特币基于 Longest-Chain 方法来增加区块链的方法会随着区块产生速率增加或者区块验证时间增加而使系统安全性降低（这里安全性降低是指，即便 attacker 没有 51%算力，仍能进行 double spend 攻击）。

2. GHOST 协议内容。

3. GHOST 协议和 Longest-Chain 方法在区块生成速率和系统安全方面的比较。

存疑：

3.1.2 中的 2D?

3.2.1 中的等式关系？

3.2.2 中的推导？

内容整理：

Part 1：Longest-Chain 规则的弊端

1.1. Bitcoin 系统的主要攻击：Double-Spend Attacks

比特币系统主要使用 6 个区块进行验证的方法，当一笔交易发出后，交易双方收到 6 个区块的确认后认为该交易已经完成（本交易的确认，和链接到该交易后的五个区块的确认）。

针对该机制，攻击者可以先发出一笔交易，接着自己秘密的制造七个合法的不含自己交易的链，当交易的接收者认为该交易合法并且将物品发送给攻击者后，该攻击者发布自己的链，因为此时自己的链是最长的链，所以该交易就从区块链中被抹掉了，也即攻击者实现了双花。

这种攻击要求攻击者有足够的运算能力，能够制造出这条不含自己交易的长链，显然当攻击者具有全网计算能力的 51%时，是可以实施该攻击的。

但是 Bitcoin 系统由于区块增长基于 Longest-Chain 规则，所以在区块产生速率增加或验证时间增加时，即便攻击者的算力不足 51%，仍能进行双花攻击。

一些符号说明：

λ：整个网络区块生成速率

λ_h：网络中诚实节点区块生成速率

$q.\lambda_h$：网络中 attacker 的区块生成速率

β：网络中区块链增长速率

b：一个 Block 的大小（单位 KB)

K：1KB 所包含的交易个数

TPS:每秒钟系统所能处理的交易个数

1.2. 系统安全：

若 $q.\lambda_h > \beta$，即$q > \frac{\beta}{\lambda_h}$，则 attacker 总可以用自己假造的子链来代替系统中已有的链。

相反，若$q.\lambda_h < \beta$，也即$q < \frac{\beta}{\lambda_h}$，则 attacker 能够伪造子链的概率随着 main chain 的增长而指数级的降低。

所以 $\frac{\beta}{\lambda_h}$可以称为系统的安全阈值。

1.3. 比特币系统吞吐量和系统安全性的关系

系统的吞吐量 TPS 与区块的大小 b ,和系统区块的生成速率 λ 有关，可以通过增加区块的大小，或者通过降低区块挖出的难度来增加区块生成的速率方法来

增加系统的吞吐量。但是因为系统的安全阈值 $\frac{\lambda_h}{\beta}$ 也与这两个参数有关，所以吞吐量增加的同时也会降低系统的安全性，也即，即便攻击者没有很多算力，仍能成功对系统进行攻击。

1.3.1 增加区块的大小：

当区块大小增加，在每个节点验证区块的时间同样会增加，也即造成区块在网络中传播时间的增加，可以想象，两个节点 A，B，A 先于 B 挖出了区块，但是 A 与 B 在网络中相隔较远，在 A 将 B 传播给 B 时，B 也挖出了区块，因而产生了分叉。A 传播到 B 时间越长，B 挖出区块的可能性更大，也即产生分叉的概率越大。

当分叉产生，相当于诚实网络中算力的分散，对于一个 hash 问题，分散的算力造成诚实节点挖出区块的概率的降低，而由于攻击者网络算力的集中，那么攻击者就会有更大的可能性先于诚实节点挖出区块，增大了系统被攻击的可能性。

1.3.2 加速区块产生：

通过降低难度可以时区块被挖出的时间降低，也即区块产生的速率增大。由于难度的降低，节点计算出 hash 的可能性就会增大，多个节点同时计算出一个问题的可能性也会增大，也即增加了系统中分叉的个数。

分叉的增多同样会使攻击者攻击难度减小。

1.4. 四者的关系如图如图所示：

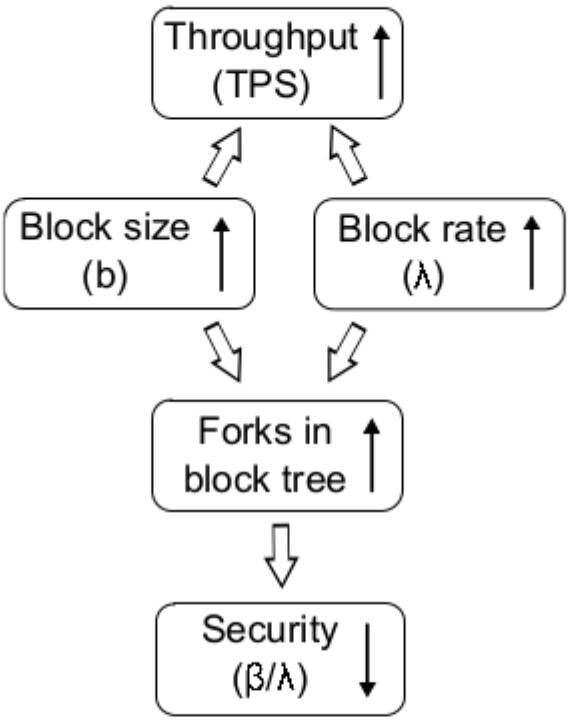


Fig. 2. A general view of tradeoffs in the Bitcoin protocol. Increasing the block size or the block rate causes an increase in the TPS, but also decreases the security from double-spend attacks.

Part2：GHOST 协议：

2.1 主要思想：

那些不被链接到主链的的区块，仍能为在进行主链选择中贡献出一部分权重。

算法如下：

Algorithm 1. *Greedy Heaviest-Observed Sub-Tree (GHOST)*
Input: Block tree T

1. *set* $B \leftarrow$ *Genesis Block*
2. *if* $Children_T(B) = \emptyset$ *then return(B) and exit*
3. *else update* $B \leftarrow \underset{C \in Children_T(B)}{\operatorname{argmax}} |subtree_T(C)|$[8]
4. *goto line 2*

其中：

$Children_T(B)$：直接连接到区块 B 上的区块集合

$subTree_T(C)$：以区块 C 为跟的子树

2.2 GHOST 协议的性质：

定义 ψ_B 为区块 B 最早确定丢失或者被系统接受的时间，则

$P_r(\psi_B < \infty) = 1$ ，也即一个区块最终要么被永久丢弃或者被永久接受。

证明如下：

$time(B)$：区块 B 不被挖出的时间

D：网络延迟的直径

t：区块 B 确定被系统接受或者丢失的时间

ε_t：表示这样一个事件，下一个区块在 t+D~t+2D 时间内被挖出，且从新区块被挖出 t+3D 之前没有新的区块被挖出。

因为系统区块广播时间为 D，所以在 t+D 时刻，网络中其他节点都已经知道区块 B 的存在，在 t+D~t+2D 时间内一个新的区块被挖出，则在 t+3D 时刻，网络中所有节点都知道新区块的存在，那么根据这个新区快的链接情况，可以确定区块 B 是被主链丢弃还是被主链接受。

因为区块挖出服从均匀分布，所以事件 ε_t 需要等待的时间是有限的，也即 t 是

一个有限的时间。

上述证明中要求在新区块挖出到 t+3D 时间内不能有其他区块被挖出是因为，若有其他区块和该区块同时被挖出，假设区块 B 和区块 C 为一个主链的两个分叉，那么若在区块 B 之后同时由挖出了两个区块 D 和 E，两者分别连接到 B 和 C 上，那么区块 B 在系统中的状态仍是不确定的。

2.3. GHOST 协议与 50%攻击

即使区块产生速率很快或者网络延迟时间很久，采用 GHOST 主链选择规则的系统的系统安全阈值 $\frac{\beta}{\lambda_h}$ 是定值 1 。

因为 Double-Spend 攻击主要是将系统已经确认在主链的区块抹去。基于 Longest-Chain 规则的系统，攻击需要的算力随着区块生成速率增加而减少，而基于 GHOST 规则系统，攻击需要的算力不会随着区块生成速率增加而减少。所以主要是在证明：一个区块被系统确认在主链上以后，随着时间的增加，其被攻击者抹去的可能性减少

命题 ：区块 B 在 $time(B)+\tau$ 时刻从主链中被删除的概率随着 τ 趋于无穷大而趋于零。

证明如下：

因为区块 B 确定被丢失或者被接受的时间为 ε_t ，所以 $\varepsilon_t > time(B)+\tau$ ，若在时刻 ε_t ，B 确定被丢弃，那么该不等式恒成立，若在时刻 ε_t ，B 被系统确认保存在主链上，那么 B 的子树增长的速率为 λ_h （诚实节点生成区块的速率），而攻击者的子树增长速率为 $q.\lambda_h$ （攻击者生成区块的速率），因为 $0<1<q$ ，所以随着时间的增加，以 B 为跟的子树增长速率大于攻击者子树增长的速率，并且是指数级的变化，因为上述命题成立。

Part 3：基于 GHOST 规则系统与基于 Longest-Chain 规则系统比较

一些符号说明：

$G=(V,E)$：是整个网路的一个子网

λ'：该子网区块生成速率

D：网络传播直径

3.1. 主链增长速率

3.1.1 基于 Longest-Chain 规则主链增长速率： $\beta \geq \frac{\lambda'}{1+\lambda' \cdot D}$

解释：因为网络生成区块的速率为λ'，所以生成一个区块所需要的时间为 $\frac{1}{\lambda'}$，因为网络传播直径为 D，所以经过 D 才能将一个区块广播到全网。

一个区块生成以后，需要经过时间 D，全网才能知道该区块的存在，接着下一个区块的生成需要时间 $\frac{1}{\lambda'}$，因为该系统基于 Longest-Chain 规则，所以一个新的区块生成以后一定会连接到最长的链上面，所以系统区块链程度增加 1 所需时间至少为 $\frac{1}{\lambda'}+D$，即系统区块链增长速率为 $\beta \geq \frac{\lambda'}{1+\lambda' \cdot D}$

3.1.2 基于 GHOST 规则主链增长速率： $\beta \geq \frac{\lambda'}{1+2\lambda' \cdot D}$

这里我能理解基于 GHOST 规则主链增加 1 所需要的时间是大于基于 Longest 规则的，但是为什么是 2D，我不太明白。

3.2. 吞吐量下界

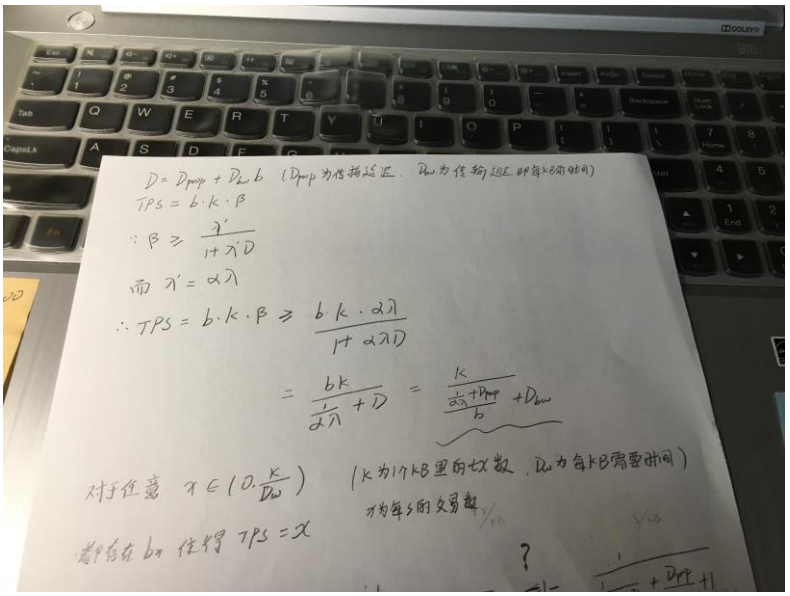
3.2.1 基于 Longest-Chain 规则的吞吐量下界，安全阈值

以下对下界的讨论均基于下面的说明：

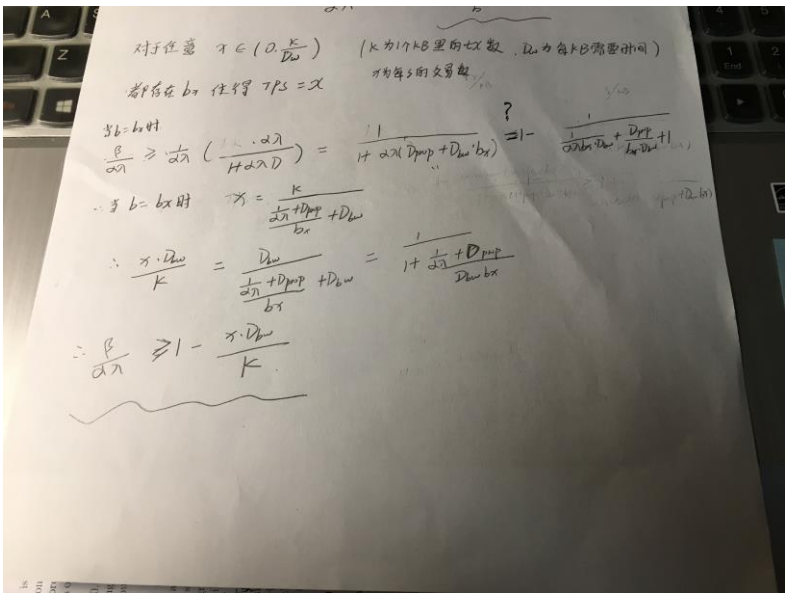
λ'：某一个诚实子网产生区块的速率

α：该子网占整个网络算力的比例，即 $\frac{\lambda'}{\lambda} = \alpha$

吞吐量下界：



安全阈值：（公式里面的等于号貌似不太对）



可以看到随着吞吐量 x 的增加，阈值在减小。

3.2.2 基于 Ghost 规则吞吐量下界，安全阈值，系统效率

吞吐量下界：因为基于 GHOST 规则的系统其区块产生速率增加不会造成其安全阈值的改变，所以其下界主要受限于网络中的带宽。

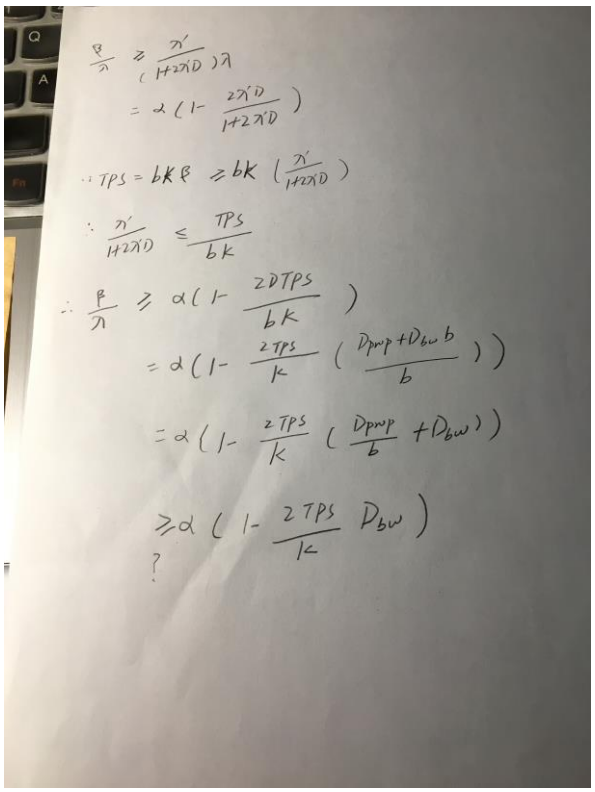
安全阈值：始终为 1

系统效率：

$\frac{\beta}{\lambda}$

这里对$\frac{\beta}{\lambda}$的分析不再是安全的角度，而是系统的效率。即该网络主链增长的速率与网络中区块产生速率的比值。

$\frac{\beta}{\lambda} = \alpha \cdot (1 - \frac{TPS \cdot 2 \cdot D_bw}{K})$推导的时候似乎缩小太多了，最后一步大于号不成立



3.3. 两者上界：

$$\beta(\lambda) \leq \frac{(p_S\lambda)^2 e^{p_S\lambda 2d} - (p_T\lambda)^2 e^{p_T\lambda 2d}}{p_S\lambda e^{p_S\lambda 2d} - p_T\lambda e^{p_T\lambda 2d}}$$

3.4 在两个系统中，区块产生速率与系统吞吐量和安全阈值之间关系：

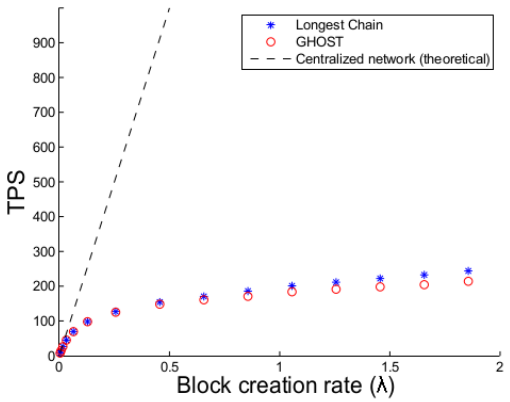


Fig. 4. $TPS(\lambda)$

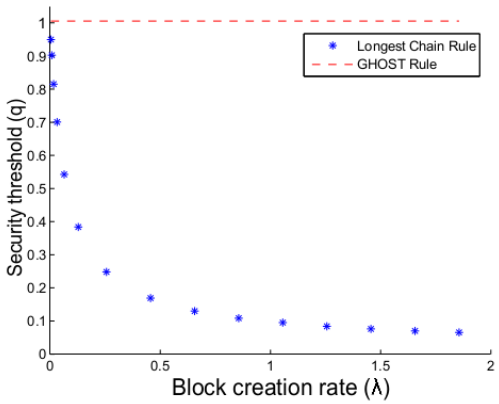


Fig. 5. $Security(\lambda)$

从 Fig 5 可以看出，随着区块产生速率增加，基于 Longest-Chain 规则的系统安全阈值在逐渐减低，而基于 GHOST 规则的系统安全阈值始终为 1，与上述推论相符。

因为在上述讨论中，基于 GHOST 规则的系统尽管区块产生速率可以在带宽允许范围内无限增加，但是由于区块产生速率的增长会带来分叉的增多，所以很多一部分区块不会被链接到主链上面，这部分区块对于系统的吞吐量是没有帮助的，但是从 Fig4 可以看出，即便存在着这样的情况，随着区块产生速率的增加，基于 GHOST 规则的系统和基于 Longest-Chain 规则的系统其系统吞吐量增加是相近的，也即分叉的增多并没有对 GHOST 系统的吞吐量造成太大的影响。

参考文献：Secure High-Rate Transaction Processing in Bitcoin