

## A Discussion of confirmation risk

In this section, we discuss how to estimate the confirmation risk when there is no active attack. “No active attack” assumes that all the blocks received by the honest nodes mined by honest nodes when estimating the confirmation risk. But the honest miners suspect that there is an attacker withholding several blocks.

Intuitively, when all the honest miners follows a block  $b$ , the block  $b$  will accumulate subtree weight faster than any sibling blocks and become irreversible with a high probability. More percisely, suppose block  $a$  is the sibling of block  $b$  and the attacker attempts to make the sibling block  $a$  have more weight than  $b$  by concentrating all its computing power in mining blocks under subtree of  $a$ . We let random variable  $X_0$  be the current subtree weight difference between tree  $a$  and tree  $b$ <sup>1</sup>, and  $X_k$  be the difference after  $k^{\text{th}}$  blocks (including the block with weight 0) added to tree  $a$  or tree  $b$ . Then  $\exists k \geq 0, X_k \geq 0$  implies subtree  $a$  has more weight than subtree  $b$  at some time.

**Estimation for  $X_0$ :**  $X_0$  represents the current subtree weight difference between tree  $a$  and tree  $b$ . All the blocks withholding by attacker are also included in computing the weight of  $a$ . But we have no way to know the exact number of these blocks. Ghost [24] assumes this number follows a poisson distribution with expectation  $\lambda_a t$ , where  $\lambda_a$  is the number of blocks generated by attacker in a unit of time and  $t$  is the time elapsed since  $b$ .parent is generated. Since all the blocks generated later than  $b$ .parent can't be in  $\text{Past}(b.\text{parent})$ , the number of blocks outside  $\text{Past}(b.\text{parent})$  could give a lower bound of  $t$ .

In the following part, we just assume the initial value  $X_0$  is known, let  $n := -X_0$  and compute the upper bound of  $\Pr[\exists k \geq 0, X_k \geq 0]$  in terms of  $n$ .

**Estimation for  $X_k$ :** By Boole's inequality, we can get

$$\Pr[\exists k \geq 0, X_k \geq 0] \leq \sum_{k \geq 0} \Pr[X_k \geq 0].$$

For each term, we can use Markov's inequality to estimate an upper bound:

$$\Pr[X_k \geq 0] \leq \min_{s > 0} E[e^{sX_k}]$$

We assume the computing power ratio between the attacker and all the participants is  $q$ , the weight  $h$  for the blocks with adaptive weight is fixed, and the initial value  $-n < 0$ . We also assume the probability distribution of  $X_{k+1}$  depends on the sequence  $X_0, \dots, X_k$ .

Given  $X_0, \dots, X_k$ , let  $q_k$  be the probability that the  $k^{\text{th}}$  block is generated by an attacker,  $\theta_k$  be the probability that the  $k^{\text{th}}$  blocks is an honest block with adaptive weight,  $\gamma_k$  be the probability that the  $k^{\text{th}}$  blocks is a malicious block with adaptive weight.  $\gamma_k, \theta_k, q_k$  should satisfy  $\gamma_k \leq q_k \leq q, \theta_k \leq 1 - q_k$ . So probability distribution of  $X_{k+1} - X_k$

$$X_{k+1} - X_k = \begin{cases} h & p = \gamma_k/h \\ 1 & p = q_k - \gamma_k \\ 0 & p = (\theta_k + \gamma_k) \cdot (h - 1)/h \\ -1 & p = 1 - q_k - \theta_k \\ -h & p = \theta_k/h \end{cases}$$

The following two lemmas estimate the upper bound of  $E[e^{s(X_{k+1}-X_k)}]$ .

**Lemma A.1.** Given  $X_0, \dots, X_k$ ,

$$E[e^{s(X_{k+1}-X_k)}] \leq \frac{q \cdot e^{sh} + (1-q) \cdot e^{-sh}}{h} + \frac{h-1}{h}.$$

**Lemma A.2.** Assume  $\theta_k = \gamma_k = 0$ , given  $X_0, \dots, X_k$ ,

$$E[e^{s(X_{k+1}-X_k)}] \leq q \cdot e^s + (1-q) \cdot e^{-s}.$$

In the following part, we use  $\psi_1(s)$  and  $\psi_h(s)$  to denote these two bounds:

$$\begin{aligned} \psi_1(s) &:= q \cdot e^s + (1-q) \cdot e^{-s} \\ \psi_h(s) &:= \frac{q \cdot e^{sh} + (1-q) \cdot e^{-sh}}{h} + \frac{h-1}{h} \end{aligned}$$

We want the attacker can not mine an block under subtree  $a$  when  $a$  is young. Here, we assume there exists  $M$  such that for all the  $k \leq M$ ,  $\theta_k = \gamma_k = 0$ . We will discuss  $M$  later.

Given  $M$ , we can get the upper bound for  $E[e^{sX_k} | X_0 = n]$ .

**Lemma A.3.** Assume  $\theta_k = \gamma_k = 0$  holds for all the  $k \leq M$ , we have

$$E[e^{sX_k} | X_0 = n] \leq e^{-sn} \cdot \psi_1(s)^{\min\{k, M\}} \cdot \psi_h(s)^{\max\{k-M, 0\}}.$$

If we use the parameter  $q = 1/11, h = 1000$ , which is consistent with our experiment, then we can get the probability for  $k \leq M$  and  $k > M$ .

For  $k \leq M$ ,

$$\begin{aligned} \Pr[X_k \geq 0] &\leq \min_s e^{-sn} \cdot \psi_1(s)^k \\ &\leq e^{-2.21n} \cdot \psi_1(2.21)^k \\ &< 0.11^n \cdot \left(\frac{13}{14}\right)^k \end{aligned}$$

<sup>1</sup>When we compute the tree weight of  $b$ , we only consider the blocks received by all the honest miners. When we compute the tree weight of  $a$ , we consider all the mined blocks, including the blocks withheld by attacker

For  $k > M$ ,

$$\begin{aligned}
\Pr[X_k \geq 0] &\leq \min_s e^{-sn} \cdot \psi_1(s)^M \cdot \psi_h(s)^{\max\{k-M, 0\}} \\
&\leq e^{-0.0022n} \cdot \psi_1(0.0022)^M \cdot \psi_h(0.0022)^{k-M} \\
&< e^{-0.0022n} \cdot (0.84)^{M/100} \cdot 0.93^{(k-M)/1000} \\
&< 0.81^{n/100} \cdot 0.84^{M/100} \cdot 0.93^{(k-M)/1000}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\Pr[\forall k, X_k \geq 0] \\
&\leq \sum_{k=0}^{+\infty} \Pr[X_k \geq 0] \\
&\leq 0.11^n \cdot \sum_{k=0}^{+\infty} (13/14)^k \\
&\quad + 0.84^{M/100} \cdot 0.81^{n/100} \cdot \sum_{k=M}^{+\infty} 0.93^{(k-M)/1000} \\
&\leq 14 \cdot 0.11^n + 0.81^{n/100} \cdot 0.84^{M/100} \cdot 13800
\end{aligned}$$

If  $M$  has a large enough lower bound with negligible exception, then  $0.84^{M/100} \cdot 13800$  will be a negligible term and  $14 \cdot 0.11^n$  becomes the dominant term.

**Estimation about M:** In the worst case,  $M = 0$ , we need  $n/100 > 106$  to make this probability less than  $10^{-6}$ . This is equivalent to eleven large weight blocks. A more tight bound shows that six large weight blocks is enough.

Fortunately, the confirmation policy doesn't work in the worst case in normal scenarios. When there is no active attack, almost all the blocks will be stable. If the attacker can mine a block with adaptive weight under subtree  $a$ , it must find a past set  $B$ , such that  $a \in \text{Pivot}(B, g)$  and

$$\begin{aligned}
&\exists a' \in \text{Pivot}(B, g), a'' = a'.\text{parent}, \\
&\text{SubStableTW}(B, a')/\text{SubTW}(B, a'') < \alpha \\
&\text{SubTW}(B, a'') > \beta
\end{aligned}$$

We let the generation time of block  $a''$  is time 0, block  $b$  is followed by all the honest nodes since time  $t_1$ , the generation time of attacker's adaptive weighted block is time  $t_2$ . Let the block generation rate of all the participants be  $\lambda$ . Also, we define some numbers

Since the network is in a normal scenario without active attack, we assume the blocks generated by honest nodes between time 0 and time  $t_1$  is in  $\text{SubStableTW}(B, a')$  except than  $c_1$  block weights. We also assume the number of honest blocks in

$$\begin{aligned}
&\text{SubTW}(B, a'') - \text{SubStableTW}(B, a') \\
&\text{has a small upper bound } c_2.
\end{aligned}$$

In exception, the attacker will mine  $q\lambda t_1$  block weights in  $[0, t_1]$  and  $q\lambda t_2$  block weights in  $[t_1, t_2]$ . Also, the honest nodes will mine  $(1-q)\lambda t_1$  in time  $[0, t_1]$  and thus  $\text{SubTW}(B, a'')$  is about  $(1-q)\lambda t_1 - c_1$

So the attacker must mine blocks with total weight  $\alpha \cdot \max\{(1-q)\lambda t_1 - c_1, \beta\} - c_2 - q\lambda t_1$  in time  $[t_1, t_2]$ , we can get an lower bound for this formula

$$(\alpha - q/(1-q)) \cdot \beta - c_2 - c_1 \cdot q/(1-q).$$

If we choose  $q = 1/11$ , and  $c_1 = c_2 = 25$ , this formula will be larger than  $(\alpha - 0.1) \cdot \beta - 30$ . Since the attacker can generate  $(\alpha - 0.1) \cdot \beta - 30$  in time interval  $[t_1, t_2]$ , the honest miner can generate  $(10\alpha - 1) \cdot \beta - 300$  blocks in the subtree of  $b$ . This implies  $n + M \geq \zeta \cdot ((10\alpha - 1) \cdot \beta - 300)$ , here  $\zeta = 0.9$  is a discount ratio to obtain a negligible exception.

When  $\alpha = 0.35$ ,  $\beta = 5000$ ,  $n + M \geq 11590$ , which will make  $0.81^{n/100} \cdot 0.84^{M/100} \cdot 13800$  smaller than  $2 \times 10^{-5}$ .

**Proof of lemma A.1 and A.2:** Let  $f_k(q_k, \theta_k, \gamma_k, s) = E[e^{s(X_{k+1}-X_k)}]$ . According to the Hölder's inequality, we have

$$\begin{aligned}
\frac{df_k}{d\gamma_k} &= \frac{e^{sh} + (h-1)}{h} - e^s \geq 0 \\
\frac{df_k}{d\theta_k} &= \frac{e^{-sh} + (h-1)}{h} - e^{-s} \geq 0
\end{aligned}$$

So  $f_k(q_k, \theta_k, \gamma_k, s) \leq f_k(q_k, 1-q_k, q_k, s)$ . Let  $g_k(q_k, s) = f_k(q_k, 1-q_k, q_k, s)$ , we have

$$\frac{dg_k}{dq_k} = \frac{e^{sh} - e^{-sh}}{h} \geq 0.$$

So  $f_k(q_k, \theta_k, \gamma_k, s) \leq g_k(q_k, s) \leq g_k(q, s)$ . It implies

$$E[e^{s(X_{k+1}-X_k)}] \leq \frac{q \cdot e^{sh} + (1-q) \cdot e^{-sh}}{h} + \frac{h-1}{h}.$$

So lemma A.1 proved.

Similarly, let  $\bar{g}_k(q_k, s) = f_k(q_k, 0, 0, s)$ , we also have  $\frac{d\bar{g}_k}{dq_k} \geq 0$ , so  $\bar{g}_k(q_k, s) \leq g_k(q_k, s)$ . It implies

$$E[e^{s(X_{k+1}-X_k)}] \leq q \cdot e^s + (1-q) \cdot e^{-s}.$$

when  $\theta_k = \gamma_k = 0$ . So lemma A.2 proved.

**Proof of lemma A.3:** According to lemma A.1 and A.2, we have that

$$E[e^{s(X_k-X_{k-1})}|X_0, \dots, X_{k-1}] \leq \begin{cases} \psi_1(s) & k \leq M \\ \psi_h(s) & k > M \end{cases}$$

If  $k \leq M$ , we have

$$\begin{aligned}
& E[e^{s(X_k - X_0)} | X_0] \\
&= E[E[e^{s(X_k - X_{k-1})} \cdot e^{s(X_{k-1} - X_0)} | X_0, \dots, X_{k-1}] | X_0] \\
&= E[e^{s(X_{k-1} - X_0)} \cdot E[e^{s(X_k - X_{k-1})} | X_0, \dots, X_{k-1}] | X_0] \\
&\leq E[e^{s(X_{k-1} - X_0)} \cdot \psi_1(s) | X_0] \\
&= \psi_1(s) \cdot E[e^{s(X_{k-1} - X_0)} | X_0]
\end{aligned}$$

Similarly, we have  $E[e^{s(X_k - X_0)} | X_0] \leq \psi_h(s) \cdot E[e^{s(X_{k-1} - X_0)} | X_0]$  for  $k > M$ .

By induction, we have

$$E[e^{s(X_k - X_0)} | X_0] \leq \psi_1(s)^{\min\{k, M\}} \cdot \psi_h(s)^{\max\{k - M, 0\}}.$$

and thus

$$E[e^{sX_k} | X_0 = -n] \leq e^{-sn} \psi_1(s)^{\min\{k, M\}} \cdot \psi_h(s)^{\max\{k - M, 0\}}.$$