Probability of Double Spend Attack for PoS Consensus with Ouroboros Praos Slot Leader Election Procedure

Lyudmila Kovalchuk[1,3], Roman Oliynykov[2,3] and Mariia Rodinko[2,3]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2] V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
[3] IOG, Ukraine
{lyudmila.kovalchuk,roman.oliynykov,mariia.rodinko}@iohk.io

Keywords: blockchain · PoS consensus · Ouroboros protocol · double spend attack

Introduction

The paper obtains upper estimates for the probability of a double spend attack on Proof-of-Stake consensus with Ouroboros Praos protocol for slot leader election. In compare with simplified models of PoS consensus, considered before in [2], we investigate two models (named Model 1 and Model 2), taking into account all features of Ouroboros Praos protocol [1]: slot creation function, active slot coefficient, possibility of empty slots and slots with multiple slot leaders (both honest and malicious), non-linear dependence between stake ratio and probability to become slot leader in some timeslot, end even non-zero synchronization time (for Model 2), measured in timeslots. In Model 1 (with zero synchronization time), we obtain explicit formulas for the exact double spend attack probability value. In Model 2 (with non-zero synchronization time), we have to make some simplification assumption in favor of the adversary. As a result, we get an upper probability estimation, using so-called dominated distribution.

The well-known double spend attack was first described in [3] and is identical for PoW- and PoS-based blockchains. According to the Double Spend (DS) attack strategy, the Adversary (malicious stakeholders) creates a transaction (which we will call *initial transaction* (ITx)) with payment to the Vendor for some goods or services.

DS Attack consists of two phases. The **first phase** of the attack lasts since the block with ITx occurs in the blockchain and till the moment when the block with initial transaction gets some definite number of *confirmation blocks* (z blocks, in our designations) and the Vendor sends the product. We assume that the Vendor sends goods or services, paid in ITx, just after he has seen the z-th confirmation block. Suppose before this moment the malicious stakeholders managed to create a longer alternative chain (more than z blocks). In that case, they share this chain with an alternative transaction just after the Vendor sends the product and gets their payment back. If they managed to create only k blocks during the first phase, for some $0 \leq k < z$, they don't share it, but continue to create new blocks in this chain, hoping to "catch up" the longest chain (created by honest stakeholders) somewhere in the future. This is the **second phase** of DS Attack, which is called "*catching up*". If malicious stakeholders succeed, they will share their longest chain and get back payment.

The main differences of the DS Attack model, described in this paper, in comparison with [2], are the next:

- existence of empty timeslots, in which no one of chains grows;
- possibility to get several slot leaders in one timeslot, who may create valid blocks, increasing several chains simultaneously (honest slot leader always add a newly created block to one of the longest chains);

- only for Model 2, we assume non-zero synchronization time for honest slot leaders.

The last feature of Model 2 needs some explanation. For Pos consensus (in contra with PoW), we call *synchronization time* the number of timeslots, needed for honest slot leaders to share the newly created block to all (or at least to all honest) nodes. We say that *synchronization time is zero* if the block created and shared by some slot leader in some timeslot becomes visible to all nodes by the end of this timeslot. Similarly, we say that *synchronization time is equal to some* $\Delta \in \mathbf{N}$ if block created and shared by some slot leader in timeslot number i is visible to all nodes not later than by the end of timeslot number $i + \Delta$.
Such assumptions mean that honest slot leaders can increase the existing longest chain only Δ timeslot later than its last block was created and shared. All blocks created and shared before this timeslot can only create a fork for the existing chain because they don't see its last block and add a newly created block to some previous one.
Our paper is organized in the next manner. Below in Section 1, we give main definitions and designations. Next, in Section 2 we provide an explicit formula (without proof, because of lack of space) for the probability of a DS attack for Model 1 and a short example of numerical results to confirm the correctness and practical usefulness of analytical results. In Section 3 we give similar results for the upper estimation of DS attack probability for Model 2 (with non-zero synchronization time). Then in Conclusion we summarize our results.

1 Main Definitions and Auxiliary Statements

In this section, we briefly describe some designations and main ideas, introduced in Ouroboros Praos protocol, and add some other necessary designations. We also formulate two statements used below in Sections 2 and 3.
Consider the blockchain with PoS consensus in which the block generation procedure is based on the Ouroboros protocol. We define $S_i, i \in I = \{1, \dots, n\} = I_H \cup I_M$, the set of all stakeholders with corresponding stake ratios α_i (here I_H / I_M are subsets of indexes corresponding to honest/malicious stakeholders).
Following the definitions and designations, introduced in [1], we define *active slot coefficient (ASC)* f and corresponding function $\varphi_f(\alpha) = 1 - (1 - f)^\alpha$, which we will call *block creation function (BCF)*, where $f \in (0, 1)$. This function defines the probability of becoming slot leader in the timeslot for the stakeholder with stake α.
Proposition 1 (properties of BCF). *In our designations BCF has the following properties for* $\alpha_i \geq 0, i = 1, 2$*:*

1. $\varphi_f(\alpha_1) + \varphi_f(\alpha_2) - \varphi_f(\alpha_1) \cdot \varphi_f(\alpha_2) = \varphi_f(\alpha_1 + \alpha_2)$;
2. $\varphi_f(\alpha_1) + \varphi_f(\alpha_2) \geq \varphi_f(\alpha_1 + \alpha_2)$.

In terms of BCF, it means that for each stakeholder S_i with stake ratio α_i the probability of being a slot leader in any timeslot is equal to $p_i = 1 - (1 - f)^{\alpha_i}$.
Note that in these assumptions the probability that the timeslot isn't empty is equal to $p = 1 - (1 - f)^1 = f$, and the probability of the inverse event (TS is empty) is equal to $1 - f$.
Define α_H total stake ratio of honest stakeholders and $\alpha_M = 1 - \alpha_H$ total stake of malicious ones.
For each timeslot (TS) $T_i, i \geq 1$, introduce the next events:

- event $H\overline{M}$ = "all slot leaders in TS T_i are honest";
- event $\overline{H}M$ = "all slot leaders in TS T_i are malicious";
- event HM = "among slot leaders in TS T_i, both honest and malicious slot leaders are present";
- event C = "TS T_i is empty (there are no slot leaders in this TS)";
- event $D = HM \cup \overline{H}M \cup H\overline{M}$ = "TS T_i is not empty (there is at least one slot leader in this TS)".

From the definition of BCF and its properties, we easily get the next Proposition.
Proposition 2. *In our designations and assumptions, the next equalities hold:*

$$P(H\overline{M}) = \varphi(\alpha_H)(1-\varphi(\alpha_M)) = (1-(1-f)^{\alpha_H})\cdot(1-f)^{\alpha_M};$$
$$P(\overline{H}M) = \varphi(\alpha_M)(1-\varphi(\alpha_H)) = (1-(1-f)^{\alpha_M})\cdot(1-f)^{\alpha_H};$$
$$P(HM) = \varphi(\alpha_H)\cdot\varphi(\alpha_M) = (1-(1-f)^{\alpha_H})\cdot(1-(1-f)^{\alpha_M});$$
$$P(C) = 1-f; P(D) = f.$$

2 Double Spend Attack Probability for Model 1

Let two different events A and B be outcomes of some experiments. Define the event $A \prec B$ as "in the following series of experiments, the event A happens earlier than B". Consider blockchain as a series of experiments, with outcomes HM, $\overline{H}M$, and $H\overline{M}$.

Proposition 3. *For events $H\overline{M}$, $\overline{H}M$, and HM, introduced above, the next equalities hold:*

$$p_H = P((H\overline{M} \prec \overline{H}M) \cap (H\overline{M} \prec HM)) = \frac{\varphi(\alpha_H)(1-\varphi(\alpha_M))}{f};$$
$$p_M = P((\overline{H}M \prec H\overline{M}) \cap (\overline{H}M \prec HM)) = \frac{\varphi(\alpha_M)(1-\varphi(\alpha_H))}{f};$$
$$p_{HM} = P((HM \prec M) \cap (HM \prec H)) = \frac{\varphi(\alpha_H)\varphi(\alpha_M)}{f}.$$

Proposition 4. *Let z be the number of confirmation blocks. Then the probabilities $P_z(k), k \in \mathrm{N} \cup \{0\}$, that malicious stakeholders create exactly k blocks during the first stage of the attack, are described with the next equalities:*

$$P_z(k) = \sum_{t=0}^{\min\{z,k\}} C_{z+k-t-1}^{z-1} \cdot C_z^t \cdot {p_M}^{k-t} \cdot {p_H}^{z-t} \cdot {p_{HM}}^t.$$

Using Propositions 2 and 3, we get the main result for Model 1 – explicit formula for DS attack probability.

Proposition 5. *In our designations, the probability P_z of success of DS attack assuming z confirmation blocks is equal to*

$$P_z = 1 - \sum_{k=0}^{z-1} P_z(k)\left(1-\left(\frac{p_M}{p_H}\right)^{z-k}\right).$$

In Table 1 below we give the number of confirmation blocks providing the probability of a double spend attack less than 10^{-3} for different stake ratios of malicious stakeholders (rows) and different values of the active slot coefficient (columns).

Table 1: The number of confirmation blocks providing the probability of a double spend attack less than 10^{-3} for different values of the active slot coefficient

p_M	f					
	0.05	**0.1**	**0.2**	**0.5**	**0.8**	**0.9**
0.05	4	4	4	4	4	4
0.1	6	6	6	6	6	7
0.2	13	13	13	13	13	14
0.3	32	32	32	32	33	35
0.4	133	133	134	135	141	149

3 Double Spend Attack Probability for Model 2

For this model, we also make the simplified assumption in favor of the adversary and assume that malicious stakeholders are well-synchronized and for them $\Delta = 0$. Introduce the next events concerning the block creation process:

- event H = "in this timeslot, one block was created by honest stakeholders and no one by malicious";
- event M = "in this timeslot, one block was created by malicious stakeholders and no one by honest".

According to our designations, the probabilities of these events may be estimated as:

$$p_H = P(H) \geq (1-f) \cdot (1-(1-f)^{\alpha_H}); \tag{1}$$

$$p_M = P(M) \leq f - (1-f) \cdot (1-(1-f)^{\alpha_H}). \tag{2}$$

The probability on the right sides of (1) and (2) form the so-called *dominant distribution* from [1]. In this distribution, we overestimate the probability of block creation by MS and underestimate the corresponding probability for HS, which causes an overestimation of DS attack probability. In what follows, we will use this dominant distribution

$$\overline{p_H} = (1-f) \cdot (1-(1-f)^{\alpha_H}),\ \overline{p_M} = f - (1-f) \cdot (1-(1-f)^{\alpha_H}), \tag{3}$$

which doesn't depend on stake distribution between all stakeholders, instead of real values p_H and p_M, which essentially depends on it. This is also one of the simplified assumptions in favor of the adversary. Note that the smaller ASC, the closer are values (3) to the left sides of (1) and (2).

Proposition 6. *Let z be the number of confirmation blocks. Then the probabilities $P_z(k), k \in \mathrm{N} \cup \{0\}$, that malicious stakeholders create exactly k blocks during the first stage of the attack, are described with the next equalities:*

$$P_z(k) = C^k_{z+k-1} \cdot \overline{p_M}^k \cdot \overline{p_H}^z.$$

Proposition 7. *In our designations, the probability P_z of success of DS attack assuming z confirmation blocks is equal to*

$$P_z = 1 - \sum_{k=0}^{z-1} P_z(k) \left(1 - \left(\frac{\overline{p_M}}{\overline{p_H}}\right)^{z-k}\right).$$

Conclusion

Our analytical results, obtained for two models, are newly, strictly proven and may be applied both to calculate the probability of a DS attack and to calculate the number of confirmation blocks, for which the attack probability is smaller than some preset bound. Recommendation, what model to choose for some particular situation, are the next. If we are sure that synchronization time is almost zero (in comparison with time slot duration), we should choose Model 1, to have an accurate value of attack probability. But if synchronization time is essential, and we want rather obtain more secure transactions than reduce confirmation time, it's better to choose Model 2.

References

[1] Bernardo David, Peter Gaži, Aggelos Kiayias, and Alexander Russell. Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain. In *Advances in Cryptology–EUROCRYPT 2018: 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29-May 3, 2018 Proceedings, Part II 37*, pages 66–98. Springer, 2018.

[2] Mikolaj Karpinski, Lyudmila Kovalchuk, Roman Kochan, Roman Oliynykov, Mariia Rodinko, and Lukasz Wieclaw. Blockchain technologies: Probability of double-spend attack on a proof-of-stake consensus. *Sensors*, 21(19):6408, 2021.

[3] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.