Extrapolation to Calculate the Probability of a Double Spending Attack

Nikolay Poluyanenko [1] [0000-0001-9386-2547], Alexandr Kuznetsov [1] [0000-0003-2331-6326], Elizaveta Lazareva [1] [0000-0002-6212-8048] and Andrey Marakushyn [2] [0000-0002-9060-5120]

[1] V. N. KarazinKharkiv National University, Svobody sq., 4, Kharkiv, 61022, Ukraine
nlfsr01@gmail.com, kuznetsov@karazin.ua,
lazareva15elizaveta@gmail.com

[2] Simon Kuznets Kharkiv National University of Economics, Nauky avenue 9a, Kharkiv, 61166, Ukraine
Andrey.marakushyn@hneu.net

Abstract. One of the important aspects of the efficiency of modern distributed networks built using blockchain technologies is the study of the security of consensus protocols. In particular, the most common cryptocurrencies and blockchain systems with probabilistic consensus protocols are subjects to so-called double-spending attack. The basis of such an attack is the use of the attacker's computing capabilities to form alternative blockchains. If the generated sequence is longer than the public chain of blocks, the attacker can present it as the proof of work and thus disrupt the correct functioning of the network. In this article we explore the probability of a successful double-spending attack, derive formulas for evaluating the corresponding events, consisting of the formation by the attacker of an alternative sequence of blocks. These formulas are extremely cumbersome and difficult to calculate. The paper proposes simplified analytical expressions to quickly assess the probability of a successful double-spending attack. For this we use the extrapolation of intermediate calculations using the Lagrange interpolation formulas, as well as binomial approximation. The simulation results show that the use of simplified expressions allows us to provide acceptable accuracy of calculations.

Keywords: blockchain technology, consensus protocols, proof-of-work, double-spending attack, polynomial interpolation, binomial approximation

1 Introduction

Modern decentralized systems are a real alternative to the classical centralized approach in creating of complex information and telecommunication systems and networks [1-4]. For example, it is advisable to use decentralization in complex systems of corporate relations, in the provision of state and administrative functions, to build independent cryptocurrencies, as well as to create secure registers, cadastres, etc. [1-

3]. These and many other aspects show the clear advantage of decentralized systems built on blockchain technology [5, 6].

At the same time, for the most developed and studied blockchain networks with probabilistic consensus protocols, there is a risk of a double-spending attack. The essence of the attack is to use the features of consensus building protocols. For example, in the proof-of-work protocols, each record in the protected registry (each block of records) can be made only by the participant who was the first to solve the exhaustive task of finding the preimage of the cryptographic hash function. The block chain stores all the important records, and each following record (block of records) is connected with the previous ones through a one-way hash function. This is a reliable and safe way to save critical data, protect it from falsification, alteration, intentional or accidental modification. At the same time, an attacker who has sufficient computing resources may try to form an alternative chain of blocks (with other records in previous blocks). This will give him the opportunity to disrupt the proper functioning of the blockchain network, for example, by spending the same assets twice (cryptocurrencies, tokens, etc.). The probability of such unauthorized interference is extremely small (with the small computing capabilities of an individual attacker it is almost impossible). However, when building a critical information infrastructure, an extremely important aspect is the assessment of various risks and threats to information security, including the probabilities of successful implementation of various attacks. Therefore, assessing the security of blockchain systems and, in particular, assessing the likelihood of a successful implementation of a double-spending attack in systems with probabilistic consensus protocols, is an urgent and practically important scientific task.

In this article, we consider consensus protocols with probabilistic completion (such as proof-of-work), and assess the probability of a successful double-spending attack. For this, we use the model of independent players (as in our previous works [4]). This model differs from the gambler's ruin model adopted in [2, 3] by both the number of elementary outcomes and general expressions for assessing the probability of a successful attack. To simplify computational calculations, we use polynomial interpolation, as well as binomial coefficients to extrapolate intermediate calculations. The results of numerical simulation show that simplified formulas can significantly reduce computational costs and obtain approximate estimates with acceptable accuracy of calculations. These results can be used to calculate the security indicators of blockchain networks with probabilistic consensus protocols.

2 The probabilities of forming a chain of blocks under the same initial conditions

To obtain the formula for the probability of a successful double-spending attack, let us consider the following possible probabilities and combinations in which the attacker succeeds:

1. Success is not possible on the first attempt ($t = N + j + k = 1 + 0 + 0 = 1$). The attacker may form no more than one block in one attempt, and in order to succeed, he needs to wait for the generation of the block by the honest network and extend the chain of blocks to one more. Thus, the attacker needs to form at least two blocks. The probability of success of the attacker is defined as: $PI_{N=1,\,j=0,k=0} = 0$

2. The attacker may succeed on the second attempt ($t = N + 0 + 1 = 2$) if he managed to form on both attempts per block, and the honest network will form only one block (no matter on the first or second attempt).The total probability of this event: $PI_{N=1,\,j=0,k=1} = p \cdot (1-p) \cdot [q \cdot q] + (1-p) \cdot p \cdot [q \cdot q] = 2 \cdot p \cdot (1-p) \cdot q^2$

3. On the third attempt ($t = N + 0 + 2 = 3$). If the honest network forms a block on the first or second attempt, then the second block should be formed by the attacker only on the third attempt, otherwise (if the second block is formed on the second attempt) there is a previous case (as for $PI_{N=1,\,j=0,k=1}$). If the honest network forms the second block on the third attempt, there are no restrictions on the formation of blocks for the attacker. The total probability of the event will be calculated similarly to the events described above. Thus, the total probability will be:

$$PI_{N=1,\,j=0,k=2} = 2 \cdot p^1 \cdot (1-p)^2 \left[2 \cdot q^2 \cdot (1-q)^1 \right] +$$
$$+ p^1 \cdot (1-p)^2 \left[3 \cdot q^2 \cdot (1-q)^1 + q^3 \right].$$

4. On the fourth attempt ($t = N + 0 + 3 = 4$) similar to the situation described in the previous paragraph, the total probability for the attacker to succeed will be equal to:

$$PI_{N=1,\,j=0,k=3} = 3 \cdot p^1 \cdot (1-p)^3 \left[3 \cdot q^2 \cdot (1-q)^2 \right] +$$
$$+ p^1 \cdot (1-p)^3 \left[6 \cdot q^2 \cdot (1-q)^2 + 4 \cdot q^3 \cdot (1-q)^1 + q^4 \right].$$

5. On the t -th attempt ($t = N + 0 + k$) the probability of success of the attacker is equal to:

$$PI_{N=1,\,j=0,k=k} = p^1 \cdot (1-p)^k \cdot$$
$$\cdot \left\{ k^2 \cdot q^2 \cdot (1-q)^{k-1} + \left[1 - \binom{k+1}{0} \cdot q^0 \cdot (1-q)^{k+1} - \binom{k+1}{1} \cdot q^1 \cdot (1-q)^k \right] \right\}$$

Carrying out similar constructions and summing all the $k = 0,1,2,\ldots$, we find the probability of a successful double-spending attack, provided that the honest network is formed of no more than $N = 1$ blocks:

$$PI_{N=1,j=0} = \sum_{k=1}^{\infty}\left\{ p\cdot(1-p)^k \cdot\left[k\cdot q^2\cdot(1-q)^{k-1} + \right.\right.$$
$$\left.\left. +\left(1-\binom{k+1}{0}\cdot q^0\cdot(1-q)^{k+1} - \binom{k+1}{1}\cdot q^1\cdot(1-q)^k\right)\right]\right\}.$$

Let us consider the case when $N=2, j=0$.

If $k=0$, as well as in the previous case, success is impossible.

If $k=1$ and, therefore, $t=N+0+1=3$, the probability of success of the attacker is equal to:

$$PI_{N=2,j=0,k=1} = 3\cdot p^2\cdot(1-p)\cdot q^3.$$

If $k=2$ ($t=4$), the probability of successful attack is equal to:

$$PI_{N=2,j=0,k=2} = 3\cdot p^2\cdot(1-p)^2\cdot\left[3\cdot q^3\cdot(1-q)\right] +$$
$$+3\cdot p^2\cdot(1-p)^2\cdot\left[4\cdot q^3\cdot(1-q)+q^4\right].$$

If $k=3$ ($t=5$), the probability of success of the attacker is equal to:

$$PI_{N=2,j=0,k=3} = 6\cdot p^2\cdot(1-p)^3\cdot\left[6\cdot q^3\cdot(1-q)^2\right] +$$
$$+4\cdot p^2\cdot(1-p)^3\cdot\left[10\cdot q^3\cdot(1-q)^2+5\cdot q^4\cdot(1-q)^1+1\cdot q^5\right].$$

The above results allow us to obtain expressions for arbitrary values of N and k:

$$PI_{N=N,j=0} = \sum_{k=1}^{\infty} p^N\cdot(1-p)^k\cdot\left\{\binom{t-1}{N}^2\cdot q^{N+1}\cdot(1-q)^{k-1} + \right.$$
$$\left. +\binom{t-1}{N-1}\cdot\left[1-\sum_{i=0}^{N}\binom{t}{i}\cdot q^i\cdot(1-q)^{t-i}\right]\right\}.$$

Let us consider the case for $N=1, j=1$.

If $k=0$ attacker's success is impossible

If $k=1$ the probability of success of the attacker is equal to:

$$PI_{N=1,j=1,k=1} = p\cdot p\cdot(1-p)\cdot\left[q\cdot q\cdot q\right].$$

If $k=2$ ($t=4$) the probability of the attacker's success is equal to:

$$PI_{N=1,j=1,k=2} = p^2\cdot(1-p)^2\cdot\left[q^3\cdot(1-q)\cdot\{3+2+2\}\right].$$

If $k=3$ ($t=5$) the probability for the attacer to succeed is equal to:

$$PI_{N=1,\,j=1,k=3} = p^{N+j}\cdot(1-p)^{k}\cdot q^{N+j+1}\cdot(1-q)^{k-1}\cdot \cdot\left\{\binom{k}{1}\cdot\binom{t-1}{2}+\binom{k-1}{1}\cdot\binom{t-2}{2}+\binom{k-2}{1}\cdot\binom{t-3}{2}\right\}.$$

Thus, summarizing the results for arbitrary initial values, we obtain the probability of a successful double-spending attack (PI) on blockchain systems using the consensus algorithm Proof of Work based on a hash function (without the attacker's advantage in one pre-formed block):

$$\begin{aligned} PI = {} & p^{N}\cdot\sum_{k=1}^{\infty}(1-p)^{k}\cdot\left\{\binom{t-1}{N}^{2}\cdot q^{(N+1)}\cdot(1-q)^{(k-1)}+\right. \\ & \left.+\binom{t-1}{N-1}\cdot\left[1-\sum_{i=0}^{N}\binom{t}{i}\cdot q^{i}\cdot(1-q)^{(t-i)}\right]\right\}+ \\ & +\sum_{j=1}^{\infty}\left\{p^{(N+j)}\cdot q^{(N+j+1)}\cdot\sum_{k=1}^{\infty}\left[sum_{p}\cdot(1-p)^{k}\cdot(1-q)^{(k-1)}\right]\right\}, \end{aligned} \tag{1}$$

where

$$t = N + j + k\,;$$

$$sum_{p} = \sum_{ip_{(N+j)}=1}^{k}\left(\sum_{ip_{(N+j-1)}=ip_{(N+j)}+1}^{k+1}\left(\cdots\sum_{ip_{2}=ip_{3}+1}^{t-2}\left(\sum_{ip_{1}=ip_{2}+1}^{t-1}\left(sum_{q}\right)\right)\right)\right);$$

$$sum_{q} =$$

$$= \sum_{iq_{(N+j)}=1}^{k}\left(\sum_{iq_{(N+j-1)}=iq_{(N+j)}+1}^{k+1}\left(\cdots\sum_{iq_{(j+1)}=iq_{(j+2)}+1}^{t-j-1}\left(\sum_{iq_{(j)}=\max\left(iq_{(j+1)}+1,ip_{(j)}\right)}^{t-j}\left(\cdots\sum_{iq_{2}=\max\left(iq_{3}+1,ip_{2}\right)}^{t-2}\left(\sum_{iq_{1}=\max\left(iq_{2}+1,ip_{1}\right)}^{t-1}(1)\right)\right)\right)\right)\right).$$

Although the expression (1) provides an accurate quantitative result on the probability of successful double-spending attacks, it also has limitations on its use, which is related to the polynomial complexity of computational calculations. In this paper, we propose using an approximate formula to calculate the probability of a successful attack. For this, simplified expressions based on polynomial interpolation can be used. In addition, we explore other approximation methods, including using binomial coefficients.

3 Extrapolation of the sum in the formula of calculation of the probability of success of the attacker

The value of sums (sum_p and sum_q) is increasing very fast and even at $j = 10$ and $k = 15$ the calculation becomes a very difficult task. For example, if $j = 20$ and $k = 7$ the sum $sum_p = 16\ 570\ 275\ 123$ (this value was calculated by the computer for several hours). On the other hand, these sums for each value N could be calculated once and used for different probabilities. Tables 1, 2, as an example, provide calculated values of sum_p for $N = 1,5$ and some values of j and k.

Table 1. Values of sum_p in the expression(1) for $N = 1$

j	k						
	1	2	3	4	5	6	7
1	1	7	25	65	140	266	462
2	1	11	58	210	602	1470	3192
3	1	16	117	563	2073	6327	16797
4	1	22	213	1314	6041	22528	71775
5	1	29	359	2761	15495	69305	260923
6	1	37	570	5345	35950	189909	833918
7	1	46	863	9690	76927	473768	2399565
8	1	56	1257	16648	154007	1093596	6327475
9	1	67	1773	27349	291592	2364642	15498742
10	1	79	2434	43256	526520	4835606	35639160
11	1	92	3265	66225	912695	9423549	77586723
12	1	106	4293	98570	1526907	17608428	161007165
13	1	121	5547	143133	2476031	31706737	320288355
14	1	137	7058	203359	3905808	55248173	613629478
15	1	154	8859	283376	6011425	93484314	1136709035
16	1	172	10985	388080	9050125	154064036	2042783757
17	1	191	13473	523225	13356092	247916850	3571657702
18	1	211	16362	695518	19357870	390392550	6090688552
19	1	232	19693	912719	27598589	602713571	10151890353
20	1	254	23509	1183746	38759285	913805304	16570275123

However, given the limited number of calculated coefficients of sum_p, the expression (1) gives a good match with the experimental data (setting up the computational experiment is given in [4]) when the probabilities q and p significantly (twice or more) differ from each other. However, there is no need to calculate the large number of values in the sum by j, that allows to be limited to some small pre-calculated set of values of sum_p. Also, at $q, p > 0,2$ the blocks will be formed with a relatively

high probability, that makes it possible to significantly reduce the sum by k. In addition, for smaller values N, the sum sum_p grows more slowly, making it possible to calculate sum_p for more values j and k.

However, in order to improve the accuracy of the calculation (increasing the calculated coefficients j and k), it is possible to extrapolate the values of sum_p using polynomial approximation. Thus, for $j = 1$ the value sum_p is very well approximated (within known values of k) and extrapolated (beyond calculated values of k) by a polynomial:

$$sum_p(N = 1, j = 1, k) = 0{,}125 \cdot k^4 + 0{,}4167 \cdot k^3 + 0{,}375 \cdot k^2 + 0{,}0833 \cdot k,$$

for $j = 2$:

$$sum_p(N = 1, j = 2, k) = 0{,}0097 \cdot k^6 + 0{,}0792 \cdot k^5 + 0{,}2431 \cdot k^4 + 0{,}3542 \cdot k^3 + \\ + 0{,}2464 \cdot k^2 + 0{,}0947 \cdot k - 0{,}2158.$$

Extrapolation is also well done by j, and the value of k is fixed.

Table 2. Values of sum_p in the expression (1) for $N = 5$.

j	k						
	1	2	3	4	5	6	7
1	1	43	631	5335	31795	148219	575107
2	1	51	900	9100	64215	350709	1578214
3	1	60	1265	15185	125925	799834	4145505
4	1	70	1745	24600	237279	1736315	10277050
5	1	81	2361	38661	429387	3587388	24053848
6	1	93	3136	59045	748230	7079128	53381664
7	1	106	4095	87850	1259860	13400268	112900788
8	1	120	5265	127660	2056860	24434838	228674250
9	1	135	6675	181615	3266253	43084995	445514295
10	1	151	8356	253486	5059063	73710049	838128214
11	1	168	10341	347755	7661745	122712954	1527675457
12	1	186	12665	469700	11369715	199311469	2705845884
13	1	205	15365	625485	16563225	316537844	4669213780
14	1	225	18480	822255	23725842	492518292	7867415920
15	1	246	22051	1068236	33465804	752091712	12969669012
16	1	268	26121	1372840	46540540	1128836172	
17	1	291	30735	1746775	63884655	1667581587	
18	1	315	35940	2202160	86641695	2427497877	
19	1	340	41785	2752645	116200021	3485859706	
20	1	366	48321	3413536	154233135	4942601727	

As established experimentally, it is desirable to use the Lagrangian interpolation polynomial for computational methods (see, for example, [7] or [8]). The essence of which is as follows. We know some values of sum_p at the first values of x_i (we assume x_i to be k_i or j_i), where $i = 1, 2, \ldots, n$, and n is the number of points by which the Lagrangian polynomial is constructed (in our case, for $N = 1$, n must be chosen as $n = 3 + 2 \cdot j$ for extrapolations by k and as $n = 2 \cdot k \pm 1$ for extrapolations by j), then we can extrapolate sum_p by Lagrange interpolation polynomial:

$$sum_p = \sum_{i=0}^{n} \left(sum_p(x_i) \cdot \prod_{\substack{s=0, \\ s \neq i}}^{n} \frac{x - x_s}{x_i - x_s} \right).$$

The results of the extrapolation of value of sum_p are illustrated in Figure 1, which shows a graph of the dependence of precisely calculated values of sum_p (solid line) by formula (1) and its approximation and extrapolation (dashed line).

Extrapolation using a polynomial gives a very good accuracy, but with increasing k (j) increases the number of first values of sum_p that must be known and whose values are taken into account in the Lagrangian interpolation polynomial. Given that we are aware of a rather limited number of sum_p (for example, in Table 2 for $k = 7$ this is only for $j \leq 15$) polynomial extrapolation will also have some limits.

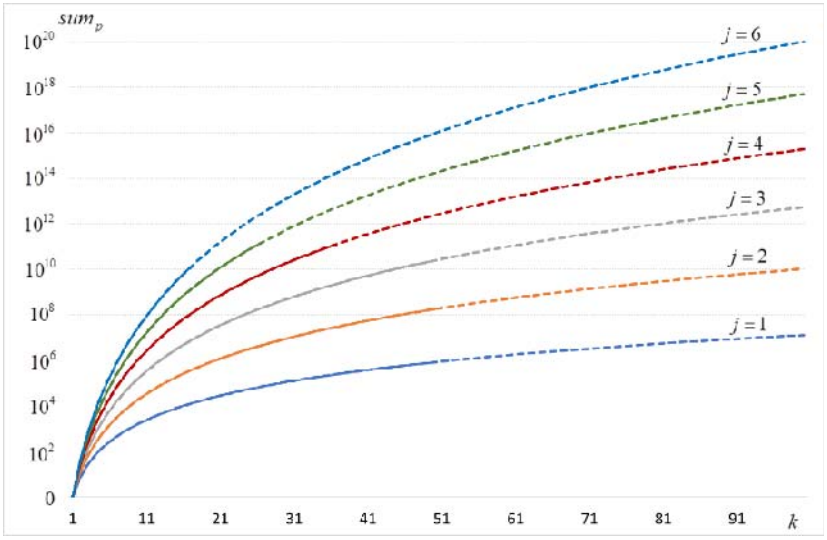
The next extrapolation method we have applied is binomial extrapolation, that is, extrapolation using binomial coefficients of the form

$$sum_p = d \cdot \binom{N + j + k_i - 1}{N + j}^2$$

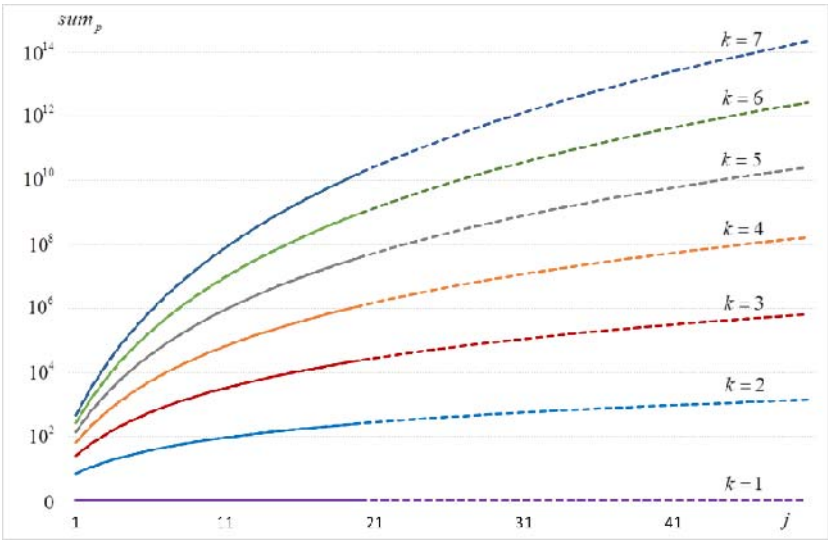
where d – some coefficient $0 < d < 1$, which is selected experimentally.

Binomial extrapolation gives much less precision than polynomial, but better than neglecting additives in general. This is especially noticeable when $q \approx p$ and a significant number of coefficients j and k must be considered.

Thus, the results obtained can significantly reduce the complexity of the calculations when calculating the probability of a successful double-spending attack. In particular, to calculate the probabilities by the formula (1), it is necessary to calculate an infinite number of sums sum_p, each of which in turn is formed by summing a large number of terms. We were able to simplify these intermediate sums by introducing interpolation polynomial equations. By extrapolating the sums to a larger data range, we can replace complex and cumbersome calculations with simple and convenient calculations based on interpolation polynomials. The simulation results show that the values calculated in this way almost 100% repeat the results obtained by exact expressions using formula (1).



а)



б)

Fig. 1. Dependencies of values of sum_p (solid line) calculated by formula (1) and its approximation and extrapolation (dashed line) for $N=1$ a) extrapolation by k b) extrapolation by j

These results can be useful for assessing the security of blockchain networks with probabilistic consensus building [9-11]. In particular, practical recommendations for constructing asset transfer protocols in decentralized systems can be justified through restrictions on the probability of corresponding risks. For example, given the restrictions on the admissible probability of loss of an asset as a result of a double-spending

attack, our formulas allow us to calculate the minimum number of confirmed transactions (length of a chain of blocks) and these calculations can be performed very quickly even in the conditions of a rapid change in the share of controlled computing resources. These and other studies are our promising areas of further work.

4 Conclusions

This paper address in detail one of the main vulnerabilities of blockchain systems built by consensus with probabilistic completeness - the double-spending attack.

Basing on the model of "independent players", we have obtained the analytical expression of the probability of a successful double-spending attack on a blockchain system which uses the consensus algorithm Proof of Work (PoW) based on a hash function depending on the number of confirmations used and the number of attempts and hash rate of both the honest network and the attacker.

The adopted model of "independent players" and the resulting formula (1) eliminates the significant disadvantages inherent in other work in this field, namely:

- the race between two participants of the network does not have to be endless, it is enough to be limited by some fixed number of attempts;
- uses a more adequate, in the authors' view, model of "independent players", which includes a space of four elementary events, instead of two, used in the model of "gambler's ruin";
- the probability of forming a block with the honest network and the attacker are independent quantities that are directly determined by the capacities possessed by the participants, and the mentionedprobabilities are independent on each other, that is, the requirement $p+q=1$ is optional;
- it is calculated the probability for the attacker to be ahead of the honest network, and not only the probability of catching up with it, that is, when the attacker does not have an advantagein one pre-formed block.

The quantitative values obtained by the expression (1) of the probability of a successful attack for different opportunities of the attacker (the probability of forming a block), the different number of formed blocks after which the agreement is considered confirmed, the different duration of the race (the number of blocks during which the attacker continues to catch up the network) are given. To simplify the numerical calculations, it is proposed to use polynomial approximation. In particular, the intermediate sums in formula (1) were able to be approximated by Lagrange interpolation polynomials. In addition, we used binomial extrapolation. However, the simulation results show that polynomial interpolation provides greater accuracy. The formulas obtained make it possible to quickly assess the probability of a successful double-spending attack using these simplified formulas.

This study can be useful for modeling and evaluating the effectiveness of blockchain networks, as well as in other practically important applications [12-16].

References

1. Zaghloul, E., Li, T., Mutka, M.W., Ren, J.: Bitcoin and Blockchain: Security and Privacy. ArXiv, abs/1904.11435 (2019)
2. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2009)
3. Rosenfeld, M.: Analysis of hashrate-based double-spending (2014)
4. Poluyanenko, N.A., Kuznetsov, A.A.: Simulation of double spend attack on the "Proof of Work" consensus protocol. Radiotekhnika. 3, 146–161 (2019). doi: 10.30837/rt.2019.3.198.11 (in Russian)
5. Zaghloul, E., Li, T., Mutka, M.W., Ren, J.: Bitcoin and Blockchain: Security and Privacy (2019)
6. Ozisik, A.P., Levine, B.N.: An Explanation of Nakamoto's Analysis of Double-spend Attacks (2017)
7. Turchak, L.I., Plotnikov, P.V.: Fundamentals of numerical methods. Moscow, Fizmat-lit. 304 p. (2003) (in Russian)
8. Interpolating Polynomial, http://dx.doi.org/10.3840/001276, (2007)
9. Pilkington, M.: Blockchain technology: principles and applications, http://dx.doi.org/10.4337/9781784717766.00019, (0)
10. Salman, T., Jain, R., Gupta, L.: Probabilistic Blockchains: A Blockchain Paradigm for Collaborative Decision-Making. In: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE (2018). doi: 10.1109/UEMCON.2018.8796512
11. Bhat, M., Vijayal, S.: A Probabilistic Analysis on Crypto-Currencies Based on Blockchain. In: 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS). IEEE (2017). doi: 10.1109/ICNGCIS.2017.37
12. Kuznetsov, A., Shekhanin, K., Kolhatin, A., Kovalchuk, D., Babenko, V., Perevozova, I.: Performance of Hash Algorithms on GPUs for Use in Blockchain. In: 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT). IEEE (2019). doi: 10.1109/ATIT49449.2019.9030442
13. Bondarenko, S., Liliya, B., Oksana, K., & Inna, G.: Modelling instruments in risk management. International Journal of Civil Engineering and Technology. 10(1), 1561-1568 (2019)
14. Runovski, K., Schmeisser, H.: On the convergence of fourier means and interpolation means. Journal of Computational Analysis and Applications. 6(3), 211-227 (2004)
15. Tkach, B.P., Urmancheva, L.B.: Numerical-analytic method for finding solutions of systems with distributed parameters and integral condition. Nonlinear Oscillations. 12, 113–122 (2009). doi:10.1007/s11072-009-0064-6
16. Chornei, R.K., Daduna V.M., H., Knopov, P.S.: Controlled Markov Fields with Finite State Space on Graphs. Stochastic Models. 21, 847–874 (2005). doi:10.1080/15326340500294520