УДК 004.056.5 **DOI:10.30837/rt.2020.1.200.12**

М.О. ПОЛУЯНЕНКО, канд. техн. наук, О.О. КУЗНЕЦОВ, д-р техн. наук

АНАЛІТИЧНЕ МОДЕЛЮВАННЯ АТАКИ ПОДВІЙНОЇ ВИТРАТИ НА БЛОКЧЕЙН-СИСТЕМИ ІЗ ЙМОВІРНІСНИМ ПРОТОКОЛОМ КОНСЕНСУСУ

1. Вступ

Чи не найголовнішим аспектом побудови безпечних та надійних розподілених децентралізованих систем за блокчейн-технологією є питання створення пов'язаних між собою безперервних ланцюжків блоків інформації, несанкціонована зміна яких унеможливлюється застосованими криптографічними механізмами. Це досягається використанням односпрямованих, стійких до колізій та пошуку прообразів криптографічних функцій, обчислені геш-значення яких від попередніх блоків включаються у наступні блоки. В результаті, несанкціоновану зміну бодай одного біту даних в попередніх блоках буде відразу виявлено, створення хибних записів, навмисне або випадкове викривлення інформаційних даних унеможливлюється. Але у разі розподіленого зберігання інформації виникає додаткова вимога синхронізації окремих ланцюгів блоків, які зберігаються різними вузлами. Ці та інші питання вирішуються шляхом застосування механізмів встановлення консенсусу, за допомогою яких після виконання певної послідовності дій безперервна послідовність блоків (блокчейн-ланцюг) стає однаковою на всіх вузлах децентралізованої мережі.

Головними завданнями при проектуванні блокчейн-систем є проведення досліджень технології, оцінка вразливостей до проведення атак, спрямованих на порушення безпеки (порушення цілісності, неспростовності, доступності та конфіденційності) інформації, що обробляється та зберігається за блокчейн-технологією. З урахуванням складності та важливості зазначених завдань актуальними є питання аналізу існуючих протоколів встановлення консенсусу в децентралізованих мережах, дослідження особливостей побудови та, безпосередньо, оцінка безпеки блокчейн-систем при застосуванні певних алгоритмів консенсусу. Важливою та актуальною є, на нашу думку, перевірка базових положень та припущень, які використовуються при моделюванні роботи блокчейн-мереж, та які безпосередньо впливають на кінцеві співвідношення стосовно оцінки ймовірності успішної реалізації певних атак на застосовані протоколи встановлення консенсусу.

В роботі досліджується одна з основних вразливості блокчейн-систем, побудованих за допомогою консенсусу з ймовірнісною завершеністю, а саме – атака подвійної витрати [1].

2. Ймовірності формування ланцюжка блоків з однаковими початковими умовами

Припустимо, що будь-який суб'єкт має у своєму розпорядженні потужності, які дають йому можливість сформувати блок за одну спробу (або за певний інтервал часу Δt) з ймовірністю p.

Зауважимо, що в умовах блокчейн-систем ймовірність p для кожного окремого суб'єкта не залежить від номера випробувань і від інших суб'єктів і визначається виключно потужністю, яку він має (справедливо для алгоритмів консенсусу Доказу виконаної роботи та його аналогів). Ймовірність p можна прив'язати до гешрейту (кількості протестованих геш-функцій в секунду), але в загальному випадку покладемо значення p – ймовірність сформувати блок за деяку умовну одиницю часу.

Ймовірність появи події A (в даному випадку – формування блоку) при кожному з нескінченної (або кінцевої, але досить великої) послідовності випробувань дорівнює p. Випадкова величина, при якій відбулася вперше подія A, є дискретною випадковою величиною. В такому випадку завдання знаходження ймовірності події A при t-му випробуванні зводиться до знаходження закону розподілу випадкової величини t.

Механізм атаки подвійної витрати докладно викладено у роботах Сатоші Накамото [2] та Мені Розенфельда [3] та багатьох інших авторів. При викладі матеріалу будемо вважати, що читач вже знайомий з цими роботами. Нагадаємо, що зловмисник може перемогти чесну мережу в момент початку гонки, тобто, коли чесною мережею сформовано N підтверджень транзакцій, стосовно якої зловмисник бажає здійснити зміни. При цьому йому необхідно сформувати N або більше блоків до того моменту коли чесна мережа сформує N блоків. Якщо йому це не вдасться, то у нього все ще є можливість наздогнати чесну мережу на $N+j$ блоці, де j – кількість блоків, сформованих чесною мережею на додаток до необхідних N блоків.

В роботах Сатоші Накамото [2] та Мені Розенфельда [3] отримано вираз з урахуванням ряду припущень, серед яких:

- ймовірність перемоги зловмисника еквівалентна задачі про «розорення гравця», тобто група подій в гонці між чесною мережею і зловмисником складається тільки з двох подій, ймовірності яких однозначно пов'язані між собою співвідношенням $p+q=1$. Ми будемо отримувати вираз ймовірності перемоги, вважаючи, що ймовірності сформувати блок чесною мережею та зловмисником є незалежними подіями, які визначаються безпосередньо потужностями, якими володіють учасники, та зазначені ймовірності ніяк не залежать один від одного, тобто будемо використовувати модель «незалежних гравців» (більш детально дивиться у [4]). Випадок, коли $p+q=1$ у моделі «незалежних гравців», є лише окремим випадком, а не обов'язковою вимогою як в моделі «розорення гравця»;

- зловмисник має можливість (бажання) нескінченно довго проводити атаку на мережу, формуючи альтернативний ланцюг, тобто зловмисник має необмежені для цього ресурси. Ми будемо виходити з більш реалістичної картини, коли зловмисник обмежується (на підставі можливості або раціональності, більш детально це розглянуто у роботах [5] та [6]) деякою максимальною кількістю спроб $t_{\max}$, що складається з N, j та кількості невдалих спроб сформувати блок чесною мережею (k), тобто $t_{\max}=N+j+k$. Якщо протягом цього часу зловмиснику не вдалося перемогти чесну мережу – йому зараховується поразка і гонка завершується. На практиці, якщо загальний гешрейт зловмисника у два рази менше чесної мережі, зловмиснику буде економічно не вигідно підтримувати процес гонки довше ніж декілька десятків блоків, тобто $n_{\max}<20$ є цілком природне (більш детально у [5]);

- вираз ймовірності успішного формування альтернативного ланцюжка зловмисником отримано для випадку, коли зловмисник лише наздожене, а не випередить, чесну мережу або, як наведено в роботах Сатоші Накамото [2] та Мені Розенфельда [3], за припущенням, що один блок був попередньо здобутий атакуючим до початку атаки, тобто не враховувалися ймовірність його формування. В даній роботі ми не будемо робити цього припущення, а отримуємо вираз безпосередньо для умови, що зловмиснику треба випередити чесну мережу, тобто сформувати хоча б на один блок більш за чесну мережу. Отримуємо ймовірність перемоги зловмисника з рівними початковими умовами, тобто коли система стартує з нульовою (з деякого початкового моменту часу) кількістю сформованих блоків як у чесній мережі так і у зловмисника.

2.1. Ймовірність перемоги зловмисника при $N=1, j=0$

Для отримання формули ймовірності успішного проведення атаки подвійної витрати зловмисником розглянемо такі можливі ймовірності і комбінації, в яких зловмисник здобуває перемогу:

1. При першій спробі ($t = N + j + k = 1 + 0 + 0 = 1$) перемога неможлива. Зловмисник може сформувати не більш одного блоку за одну спробу, а для перемоги йому необхідно дочекатися формування чесною мережею блоку і поширити ланцюжок блоків на один більше. Таким чином, зловмиснику необхідно сформувати як мінімум два блоки.

Ймовірність перемоги зловмисника буде визначатися як: $PI_{N=1,\,j=0,k=0} = 0$;

2. При другій спробі ($t = N + 0 + 1 = 2$) зловмисник може перемогти, якщо йому вдасться сформувати за обидві спроби по блоку, а чесна мережа сформує тільки один блок (неважливо за першу або другу спробу).

Загальна ймовірність даної події:

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

3. При третій спробі ($t = N + 0 + 2 = 3$): якщо чесна мережа сформує блок при першій або другій спробі, то другий блок повинен бути сформований зловмисником тільки на третій спробі, в іншому випадку (якщо другий блок буде сформований на другій спробі) буде попередній випадок (як для $PI_{N=1,\,j=0,k=1}$). Якщо чесна мережа формує другий блок на третій спробі, то на зловмисника не накладаються жодних обмежень з можливості формування блоків.

Сумарна ймовірність настання подій буде обчислюватися аналогічно описаним подіям, підсумкова ймовірність якої буде:

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

4. При четвертій спробі ($t = N + 0 + 3 = 4$) – аналогічно ситуації, описаної в попередньому пункті. Сумарна ймовірність перемоги зловмисника буде визначатися:

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

5. При t-й спробі ($t = N + 0 + k$) ймовірність перемоги зловмисника визначається:

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

Проводячи спрощення, розкладаючи вираз $(a+b)^n$ в степеневий ряд, отримуємо вираз в наступному вигляді:

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

Проводячи аналогічні побудови і підсумовуючи всі $k=0,1,2,\ldots$ ми знайдемо ймовірність успішного проведення зловмисником атаки подвійної витрати за умови, що чесною мережею сформовано не більше $N=1$ блоків:

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

Зауважимо, що $\binom{k+1}{0}=1$ и $\binom{k+1}{1}=k+1$, однак тут ми залишаємо саме в такому вигляді для можливості подальшого більш наочного узагальнення виразу.

2.2. Ймовірність перемоги зловмисника при довільному N та $j=0$

Розглянемо випадок для $N=2,\ j=0$.

При $k=0$, як і в попередньому випадку, перемога неможлива.

При $k=1$ і, отже, $t=N+0+1=3$, ймовірність перемоги зловмисника буде визначатися:

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

При $k=2$ ($t=4$), ймовірність перемоги зловмисника визначається:

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

При $k=3$ ($t=5$) ймовірність перемоги зловмисника визначається:

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

Наведені результати дають можливість отримати вирази для довільного значення N і k:

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

Спрощуючи наведений вираз, отримуємо ймовірність перемоги зловмисника для довільного значення N і k, але за умови, що чесна мережа сформувала не більше N блоків ($j=0$):

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

2.3. Ймовірність перемоги зловмисника при формуванні чесною мережею більш N блоків ($j>0$)

Розглянемо випадок для $N=1,\ j=1$.

При $k=0$ перемога зловмисника неможлива.

При $k=1$ ймовірність перемоги зловмисника визначається:

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

При $k=2$ ($t=4$) ймовірність перемоги зловмисника визначається:

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

При $k=3$ ($t=5$) ймовірність перемоги зловмисника визначається:

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

Таким чином, узагальнюючи наведені результати для довільних початкових значень, отримуємо ймовірність успішного проведення зловмисником атаки подвійний витрати (PI) на блокчейн-системи, що використовують алгоритм консенсусу Доказ виконаної роботи на основі геш-функції (без фори зловмисника в один попередньо сформований блок):

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

де $t=N+j+k$;

$$sum_p=\sum_{ip_{(N+j)}=1}^{k}\left(\sum_{ip_{(N+j-1)}=ip_{(N+j)}+1}^{k+1}\left(\dots\sum_{ip_2=ip_3+1}^{t-2}\left(\sum_{ip_1=ip_2+1}^{t-1}\left(sum_q\right)\right)\right)\right);$$

$$sum_q=\sum_{iq_{(N+j)}=1}^{k}\left(\sum_{iq_{(N+j-1)}=iq_{(N+j)}+1}^{k+1}\left(\dots\sum_{iq_{(j+1)}=iq_{(j+2)}+1}^{t-j-1}\left(\sum_{iq_{(j)}=\max\left(iq_{(j+1)}+1,ip_{(j)}\right)}^{t-j}\left(\dots\sum_{iq_2=\max(iq_3+1,ip_2)}^{t-2}\left(\sum_{iq_1=\max(iq_2+1,ip_1)}^{t-1}(1)\right)\right)\right)\right)\right).$$

3. Екстраполяція суми у формулі розрахунку ймовірності перемоги зловмисника

Незважаючи на те, що вираз (1) надає можливість отримати точний кількісний результат щодо ймовірності атак подвійної витрати, він також має і обмеження стосовно можливості його застосування, що пов'язано з поліноміальною складність обчислювальних розрахунків.

Значення сум (sum_p та sum_q) дуже швидко зростає і вже при $j=10$ і $k=15$ розрахунок стає обчислювальна дуже складною задачею. Наприклад, при $j=20$ та $k=7$ сума $sum_p=16\ 570\ 275\ 123$ (це значення комп'ютер розраховував декілька годин). З іншого боку, ці суми для кожного значення N можна підрахувати однократно і використовувати їх для

різних ймовірностей. У табл. 1, 2, в якості прикладу наведено розраховані значення sum_p для $N = 1,5$ та деяких значень j та k.

Таблиця 1

Значення sum_p у виразі (1) для $N = 1$

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

Таблиця 2

Значення sum_p у виразі (1) для $N = 5$

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	

Однак, враховуючи обмежену кількість обчислених коефіцієнтів sum_p, вираз (1) дає хороший збіг з експериментальними даними (постановка обчислювального експерименту докладно наведено у [4]), коли ймовірності q та p значно (у два і більше разів) відрізняються один від одного. При цьому відсутня необхідність прораховувати велику кількість значень у сумі по j, що дозволяє обмежитись деякою невеликою попередньо обчисленою множеною значень sum_p. Також, при $q, p > 0,2$ блоки будуть формуватися з відносно високою ймовірністю, що дає можливість значно скоротити суму по k. Крім того, при менших значеннях N сума sum_p зростає повільніше, що дає змогу прорахувати sum_p для більшої кількості значень j та k.

Однак для підвищення точності обчислення (збільшення урахованих коефіцієнтів j та k) є можливість провести екстраполяцію значень sum_p за допомогою поліноміальної апроксимації. Так, для $j=1$ значення sum_p дуже добре апроксимується (в межах відомих значень k) та екстраполюється (за межами вже обчислених значень k) поліномом

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

для $j=2$:

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

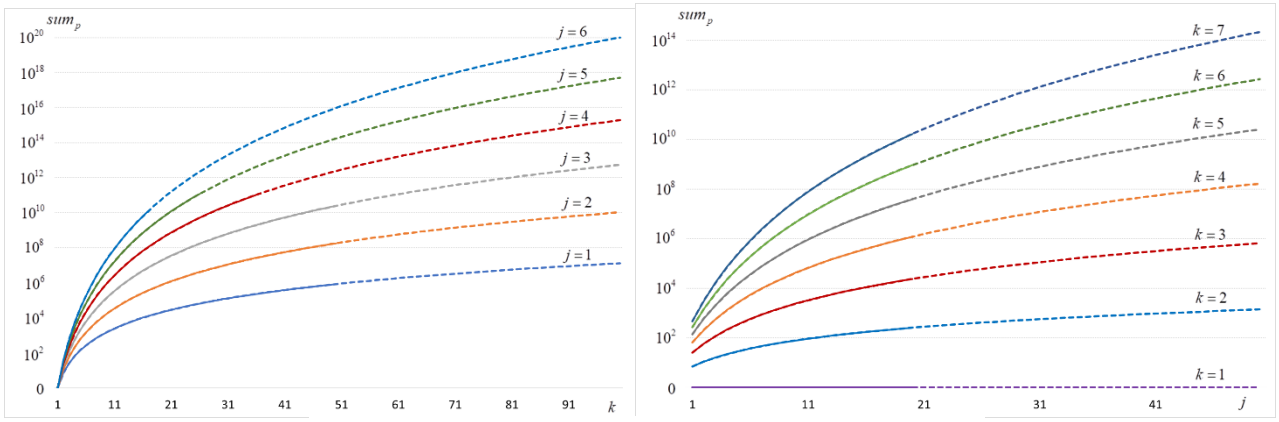
Екстраполяція також добре здійснюється по j, при цьому фіксується значення k.

Як встановлено дослідним шляхом, при обчислювальних методах бажано використовувати інтерполяційний многочлен Лагранжа (дивиться, наприклад, [7] або [8]). Суть якого полягає в наступному. Нам відомі деякі значення sum_p при перших значеннях x_i (x_i вважаємо k_i або j_i), де $i=1,2,\ldots,n$, а n – кількість точок, за допомогою яких будується многочлен Лагранжа (в нашому випадку, для $N=1$, n повинно обиратися як $n=3+2\cdot j$ для екстраполяцій по k та $n=2\cdot k\pm 1$ – для екстраполяцій по j), тоді sum_p можна буде екстраполювати за допомогою інтерполяційного многочлена Лагранжа:

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

Результати екстраполяції значення sum_p ілюструє рис. 1, де наведено графік залежності точно обчислених значень sum_p (суцільна лінія) за формулою (1) та її апроксимації та екстраполяції (пунктирна лінія).

Екстраполяція за допомогою полінома дає дуже добру точність, але зі зростанням k (j) зростає кількість перших значень sum_p, які повинні бути відомими та значення яких враховуються у інтерполяційному многочлені Лагранжа. Враховуючи те, що нам відома досить обмежена кількість sum_p (наприклад, у табл. 2 для $k=7$ це sum_p лише для $j\le 15$) поліноміальна екстраполяція також буде мати свої межі застосування.



а *б*

Рис. 1. Залежності значень sum_p (суцільна лінія) обчислених за формулою (1) та її апроксимації та екстраполяції (пунктирна лінія) для $N=1$: *а* – экстраполяция по k, *б* – экстраполяция по j

Наступним методом екстраполяції, який ми застосували, є біноміальна екстраполяція, тобто екстраполяція за допомогою біноміальних коефіцієнтів виду

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

де d – деякий коефіцієнт $0 < d < 1$, підбирається дослідним шляхом.

Біноміальна екстраполяція дає значно меншу точність (див. рис. 3, 6, 9), ніж поліноміальна, але кращу ніж неврахування доданків взагалі. Особливо це помітно, коли $q \approx p$ та треба враховувати значну кількість коефіцієнтів j та k.

4. Кількісні розрахунки. Ймовірність вдалого розгалуження блокчейн-ланцюга

Вираз (1) дає можливість наблизити обчислення до реальних умов та не застосовує ряд обмежень та припущень, які притаманні існуючим найбільш популярним аналогічним виразам, що вигідно відрізняє запропоновану модель від існуючих.

При врахуванні фіксованого числа спроб ($t_{\max} = N + n_{\max} + k_{\max}$, де $n_{\max}$ – кількість вдалих, а $k_{\max}$ – невдалих спроб чесної мережі) зловмисником наздогнати блокчейн-ланцюг чесної мережі вираз (1) з напівнескінченного ряду приймає вигляд

$$PI = p^N \cdot \sum_{k=1}^{k_{\max}} (1-p)^k \cdot \left\{ \binom{t-1}{N}^2 \cdot q^{(N+1)} \cdot (1-q)^{(k-1)} + \binom{t-1}{N-1} \cdot \left[1 - \sum_{i=0}^{N} \binom{t}{i} \cdot q^i \cdot (1-q)^{(t-i)} \right] \right\} +$$

$$+ \sum_{j=1}^{n_{\max}} \left\{ p^{(N+j)} \cdot q^{(N+j+1)} \cdot \sum_{k=1}^{k_{\max}} \left[sum_p \cdot (1-p)^k \cdot (1-q)^{(k-1)} \right] \right\},$$

значення sum_p залишається таким же як у (1).

Обчислені за даним виразом графіки ймовірності вдалого формування зловмисником альтернативного блокчейн-ланцюга, для довільного значення p та q наведені на рис. 2, 5, 8. При цьому, для $N=1$ було обрано $n_{\max}=50, k_{\max}=300$; для $N=3$ – $n_{\max}=20, k_{\max}=300$; для $N=5$ – $n_{\max}=20, k_{\max}=300$. На рис. 3, 6, 9 та 4, 7, 10 наведені відносна похибка отриманих розрахунків та кількість реалізацій моделі тестування відповідно. Тестування проводилось як описано у [4]).

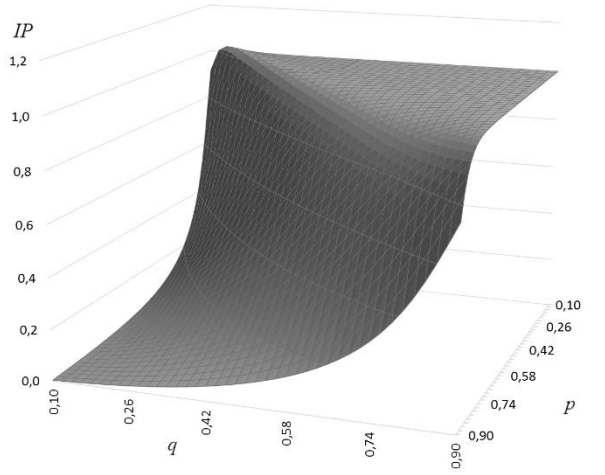


Рис. 2. Функція розподілу ймовірності для різних значень p та q при довжині сформованого ланцюжка з $N=1$ блок

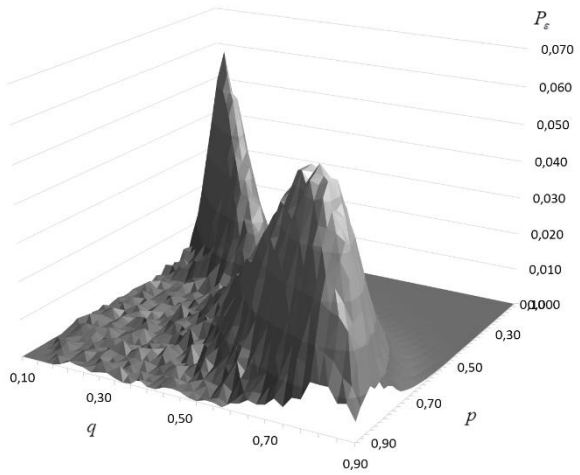


Рис. 3. Відносна похибка розбіжностей між експериментальними і теоретичними значеннями при $N=1$ блок

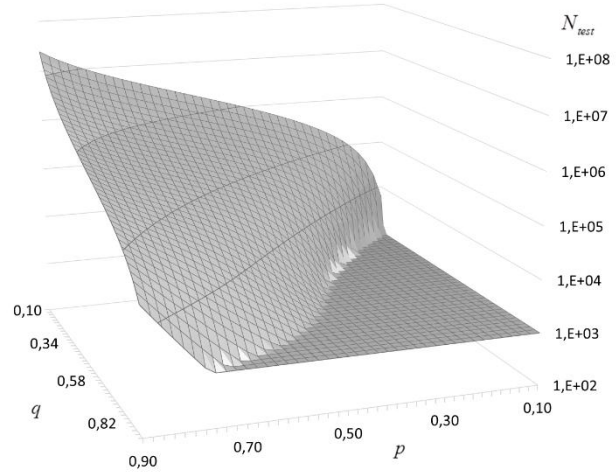


Рис. 4. Число проведених випробувань, що використовувались у експериментальних обчисленнях при довжині сформованого ланцюжка з $N=1$ блок

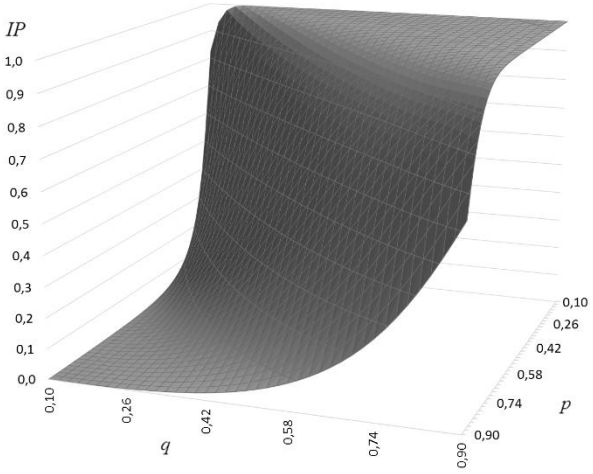


Рис. 5. Функція розподілу ймовірності для різних значень p та q при довжині сформованого ланцюжка з $N=3$ блоки

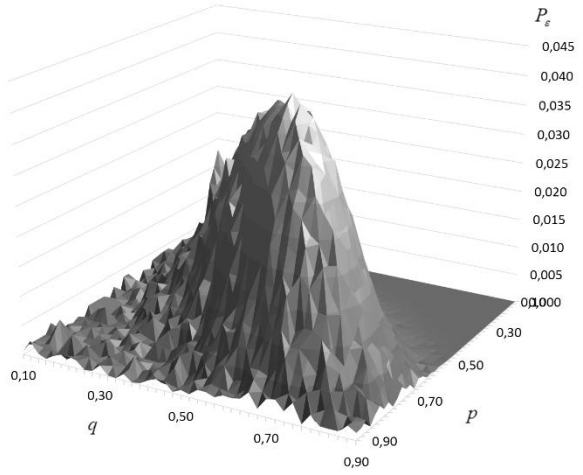


Рис. 6. Відносна похибка розбіжностей між експериментальними і теоретичними значеннями при $N=3$ блоки

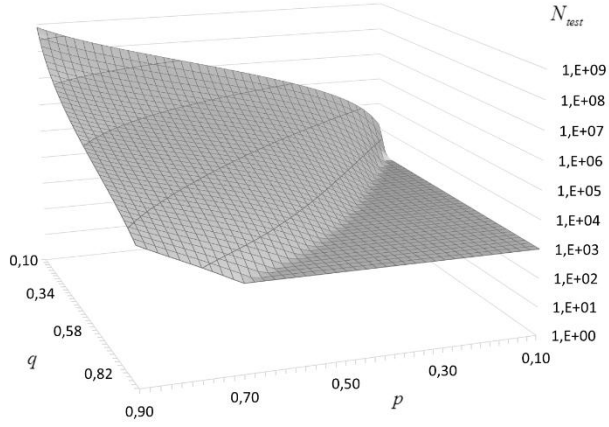


Рис. 7. Число проведених випробувань, що використовувались у експериментальних обчисленнях при довжині сформованого ланцюжка з $N=3$ блоки

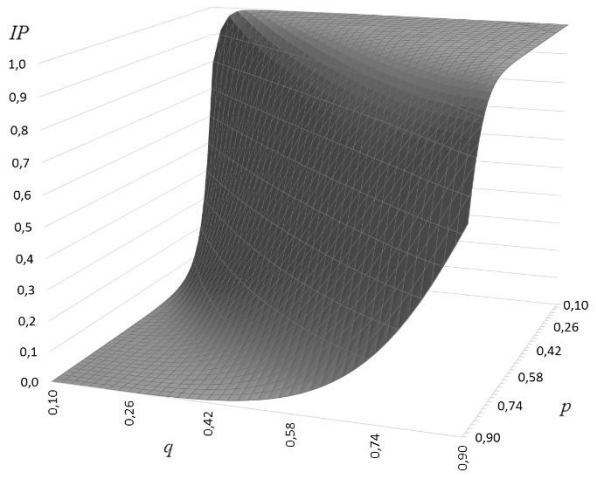


Рис. 8. Функція розподілу ймовірності для різних значень p та q при довжині сформованого ланцюжка з $N=5$ блоків

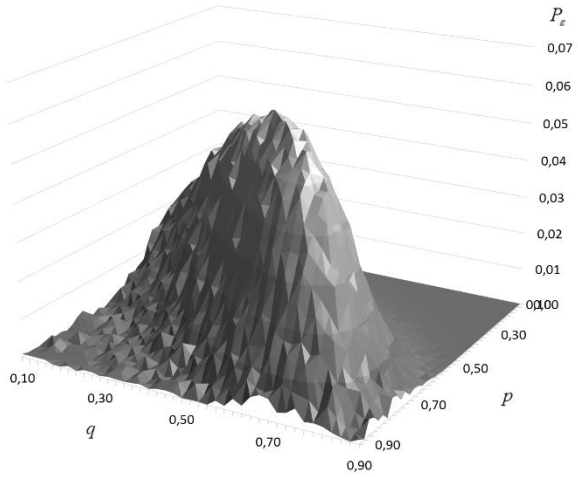


Рис. 9. Відносна похибка розбіжностей між експериментальними і теоретичними значеннями при $N=5$ блоків

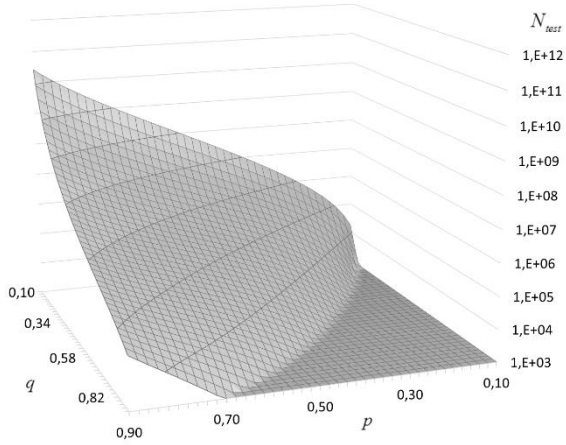


Рис. 10. Число проведених випробувань, що використовувались в експериментальних обчисленнях при довжині сформованого ланцюжка з $N=5$ блоків

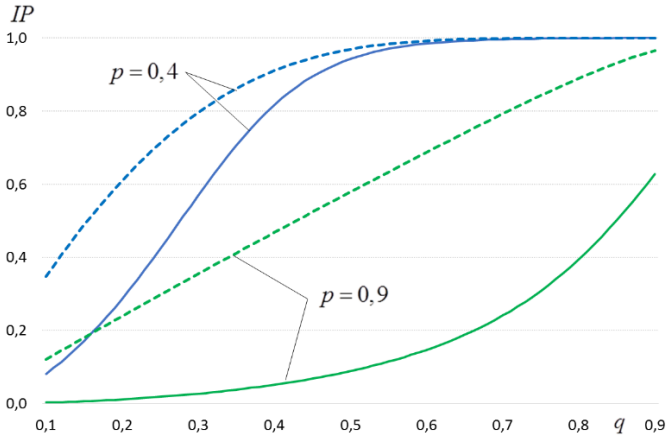


Рис. 11. Функція розподілу ймовірності при $N=1$. Пунктирна лінія – ймовірність з одним попередньо сформованим блоком зловмисником, суцільна – ймовірність з однаковими початковими умовами

Як бачимо, при $q \approx p$ відносна похибка вище обраного значення (обрано відносну помилку не гірше 1 % при довірчій ймовірності не менш 99 %), що обумовлено значно нижчою точністю екстраполяції, яку забезпечує біноміальна екстраполяція.

На рис. 11 наведене порівняння ймовірностей вдалого розгалуження блокчейн-ланцюга за умови одного попередньо сформованого блоку зловмисником (розраховано за тими ж умовами, крім попередньо сформованого блоку, що і вираз (1) – пунктирні лінії) та з однаковими початковими умовами (розраховано за формулою (1) – суцільні лінії), при $n_{\max}=50, k_{\max}=300$.

Як бачимо, за умови відсутності фори, ймовірність перемоги зловмисника суттєво знижується, що є природно, у зв'язку з тим, що зловмиснику необхідно випередити чесну мережу, тобто сформувати на один блок більше. При цьому, як бачимо, різниця є досить значна та може досягати кілька разів. Таким чином, при рівних ймовірностях зловмиснику потрібні значно більші потужності (гешрейт), ніж потужності, розраховані за умови заздалегідь сформованого зловмисником блоку.

5. Висновки

Докладно розглянута одна з основних вразливостей блокчейн-систем, побудованих за допомогою консенсусу з вірогідною завершеністю – атака подвійної витрати.

На підставі моделі «незалежних гравців» отримано аналітичний вираз розрахунку ймовірності успішного проведення зловмисником атаки подвійної витрати на блокчейн-системи, що використовують алгоритм консенсусу Доказ виконаної роботи (PoW) на основі геш-функції у залежності від використовуваної кількості підтверджень та кількості спроб, а також гешрейту чесної мережі і зловмисника.

Прийнята модель «незалежних гравців» та отримана на її основі формула (1) дозволяють позбавитись значних недоліків, які притаманні іншим роботам в даній галузі, а саме:

– проведення гонки між двома учасниками мережі не потрібно уявляти нескінченою, достатньо обмежуватись деяким фіксованим числом спроб;

– використовує більш адекватну, на погляд авторів, модель незалежних гравців, яка включає в себе простір з чотирьох елементарних подій, замість двох, що використовується у моделі «розорення гравця»;

– ймовірності сформувати блок чесною мережею і зловмисником є незалежними величинами, які визначаються безпосередньо потужностями, якими володіють учасники, та зазначені ймовірності ніяк не залежать один від одного, тобто вимога $p+q=1$ є не обов'язковою;

– обчислюється ймовірність саме випередження зловмисником чесної мережі, а ні тільки ймовірність її наздогнати, тобто коли зловмисник не має фори в один попередньо сформований блок.

Наведено кількісні значення, отримані за виразом (1), ймовірності вдалої атаки для різних можливостей зловмисника (ймовірності сформувати блок), різної кількості сформованих блоків, після яких угода вважається підтвердженою, різної тривалості гонки (кількості блоків, протягом яких зловмисник продовжує спроби наздогнати чесну мережу). За допомогою комп'ютерного моделювання експериментально перевірено розраховані за формулою (1) значення. Всі емпіричні оцінки отримано для високої точності (відносна помилка не гірше 1 %) і достовірності (довірча ймовірність не менш 99 %). Для підтвердження адекватності отриманих результатів наведено порівняння емпіричних результатів з теоретичними розрахунками.

Дослідження дозволили отримати нові аналітичні оцінки ймовірностей реалізації атак подвійної витрати на блокчейн-системи із протоколом консенсусу Доказу виконаної роботи. Ці аналітичні оцінки є відмінними від отриманих та відомих раніше, оскільки побудовані на іншій системі припущень та базових положень стосовно моделювання дій різних гравців в блокчейн-мережах, що застосовують Доказ виконаної роботи.

На основі отриманих результатів можна зробити висновок, що безпека блокчейн-систем, які використовують алгоритми консенсусу з ймовірнісною завершеністю (доказ виконаної роботи та її модифікації), мають більш високу надійність, ніж вважалось раніше.

Отримані результати можуть бути корисними при обґрунтуванні конкретних показників та параметрів протоколу консенсусу на основі Доказу виконаної роботи, при застосуванні його у якості основного механізму встановлення консенсусу перспективних децентралізованих розподілених систем та мереж, побудованих за технологією блокчейн.

Список літератури:

1. Zaghloul E., Li T., Mutka M.W. & Ren, J. (2019). Bitcoin and Blockchain: Security and Privacy. ArXiv, abs/1904.11435

2. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System / Satoshi Nakamoto, 2009. 9 с.

3. Rosenfeld M. Analysis of hashrate-based double-spending / Meni Rosenfeld, 2014. 13 с.

4. Полуяненко Н.А., Кузнецов А.А. Моделирование атаки двойной траты на протокол консенсуса «proof of work» // Радиотехника. 2019. № 198. С. 146–161. DOI: 10.30837/rt.2019.3.198.11

5. Zaghloul E., Li T., Mutka M.W. & Ren J. (2019). Bitcoin and Blockchain: Security and Privacy. ArXiv, abs/1904.11435

6. A. Pinar Ozisik, Brian Neil Levine. An Explanation of Nakamoto's Analysis of Double-spend Attacks https://arxiv.org/pdf/1701.03977.pdf

7. Турчак Л.И., Плотников П.В. Основы численных методов : учеб. пособие ; 2-е изд., перераб. и доп. Москва : Физматлит, 2003. 304 с.

8. Archer, Branden and Weisstein, Eric W. Lagrange Interpolating Polynomial. From MathWorld-A Wolfram Web Resource. http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html.

Харківський національний
університет імені В. Н. Каразіна;
АТ «Інститут інформаційних технологій»

Надійшла до редколегії 05.02.2020