Оценка эффективности Fork-атаки на протокол “блокчейн”

М.А. Черепнёв

***Аннотация*---В работе рассмотрен протокол защищенного формирования и хранения базы данных “блокчейн”. Этот протокол может быть рассмотрен как некоторое развитие протоколов коллективной электронно-цифровой подписи и контрольных сумм на случай, когда ответственность за содержание базы данных лежит на всех участниках обмена. Построена математическая модель работы протокола “блокчейн”. Мы предполагаем, что так называемое «условие длинных цепочек» является неотъемлемой частью этого протокола. Кроме того предполагается, что в случае обнаружения ошибки, содержащий её блок удаляется вместе со всеми идущими после него блоками. В рамках этой модели получены оценки на вероятность исправления ошибок в цепочке блоков, а также на скорость роста цепи и показано, что изменения этих оценок в случае, когда часть абонентов продвигают ошибочный блок, могут быть минимизированы выбором параметров безопасности. А именно, если увеличить среднее число проверяемых блоков в каждом шаге построения текущего блока «честным» абонентом, то любое фиксированное количество бракованных блоков может быть нейтрализовано при помощи правила длинных цепочек. Правда, скорость развития дерева блокчейна может стать отрицательной, то есть схема не будет работать. В статье представлены количественные оценки.**

***Ключевые слова*---Блокчейн, Fork-атака, скорость развития дерева блокчейна.**

I. Введение

Технология “блокчейн” [1] решает задачу децентрализованного защищенного формирования базы данных. Это может быть использовано не только в платежных системах, но и для хранения юридически значимых документов при построении систем электронного документооборота на предприятиях, в государственной сфере или в организациях, осуществляющих межгосударственный обмен.

Для проведения нашего анализа будем пользоваться следующей математической моделью. Пусть имеется коммуникационная сеть, обеспечивающая обновление, хранящейся у каждого абонента копии одной базы данных, которое можно считать одновременным.

Статья получена 8 февраля 2019.
Работа поддержана грантом РФФИ 18–29–03124 мк.
М.А.Черепнев - Московский государственный университет имени М.В.Ломоносова, РФ (e-mail: cherepniov@gmail.com)

Пусть база данных состоит из двух частей: защищенного реестра, который строится при помощи технологии блокчейн и некоторого набора записей, ещё не занесенных в реестр.

Реестр организован в виде электронной бухгалтерской книги, обеспечивающей защищенное и безопасное хранение записей из базы, удовлетворяющих некоторому набору формальных требований (например, это платежные поручения, содержащие электронные подписи участников обмена). Поскольку записи в реестр могут делать все участники, то имеется механизм консенсуса, позволяющий вносить исправления в реестр в случае, если с этими исправлениями согласно “большинство” участников обмена. Понятие “большинства” можно варьировать предоставлением или ограничением прав участникам обмена, однако имеется общая составляющая всех, известных на сегодня, видов консенсуса - правило “длинных цепочек”, которое должны соблюдать все “честные” участники обмена. Это правило, по - существу, и отличает старые технологии обеспечения целостности информации, основанные на контрольных суммах и электронно-цифровых подписях [2], от технологии “блокчейн”. На правиле длинных цепочек основаны практически все механизмы консенсуса, применяемые в блокчейне (POW, POS). Отметим также ещё одно естественное правило, которое должны соблюдать все “честные” участники обмена – если найдена ошибка в каком-то блоке, то все блоки, стоящие в цепочке за ошибочным блоком признаются недействительными и подлежат расформированию.

II. Определение блокчейна

Дерево блокчейна состоит из блоков (совокупностей записей). Ветвь дерева блокчейна, вообще говоря, может начинаться с любого его блока. Блоки в дереве связаны при помощи вычисления хеш функций в соответствии со следующей процедурой:

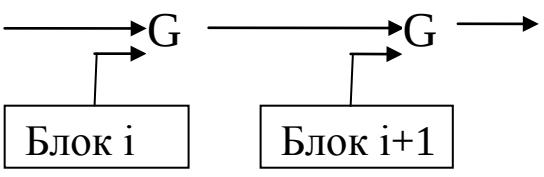

Рис. 1 Дерево блокчейна

Процедура G состоит из следующих действий:

1.Выбор наибольшей ветви дерева блокчейна и её последнего блока.

2.Проверка целостности выбранной ветви и содержания нового блока, вычисление хеш значения h этого содержания.

3.Вычисление хеш функции, которая имеет два входа (h и результат применения предыдущей процедуры в данной цепи) и один выход.

4.

Хеш функция, используемая на втором этапе, обычно та же, что и на третьем (SHA 256 для Bitcoin) и применяется несколько раз, чтобы результат существенно зависел от всего содержания блока.

III. Формальная модель и формулировки задач.

Пусть имеется формально бесконечная цепочка блоков блокчейна с номерами $\dots,-2,-1$. Пусть каждый из них может быть бракованным с небольшой, но одинаковой вероятностью q. Конечно на практике это условие, скорее всего, не выполнено. Например, если бракованный блок уже проверялся на предыдущем шаге, то он, скорее всего, был исправлен. То есть, вместо него выпущен другой блок, который также может быть бракованным с вероятностью q. Однако, если блок прошёл проверку успешно, то на следующих шагах он не будет бракованным с вероятностью 1. Если расстановка вероятностей проверки на следующих шагах не приведёт к тому, что он не будет больше проверяться никогда, то это может повлиять на вероятности исследуемых в этой статье событий. Здесь мы эту возможность не рассматриваем и считаем, что работа с деревом блокчейна организована так, что вероятность того, что блок дерева является бракованным, не зависит от номера этого блока в дереве, который в нашей модели убывает с ростом основной ветви дерева блокчейна.

При создании очередного блока (с номером 0) блок с номером $-i$ проверяется с вероятностью p_i. Будем считать, что на каждом шаге проверяется в среднем k

$$\sum_{i=1}^{N} p_i = k, p_i = 0$$

блоков, то есть , при $i > N$. Отдельно будем считать вероятности для случая, когда a процентов участников сети перестают проверять блоки (Fork - атака) [3]. Рассмотрим следующие задачи для представленной модели

А) Найти вероятность того, что хотя бы один брак будет установлен на шаге с номером $-i$.

Б) Найти вероятность того, что хотя бы один брак будет установлен на шаге с номером $j, j = 1,2,\dots$, и не будет установлен на шагах с номерами 0,1,…, *j*-1.

В) Найти матожидание и дисперсию номера шага, на котором впервые будет обнаружен хотя бы один брак.

Г) Найти скорость роста цепи блокчейна (за единицу времени взять время создания одного блока). То есть найти среднее число добавленных блоков к основной цепи дерева блокчейна в единицу времени, учитывая, что блоки могут не только добавляться, но и отбрасываться.

Д) Найти вероятность того, что бракованный блок так и останется неисправленным.

IV. Решения задач.

А) Вероятность обнаружить ошибку в блоке с номером $-i$ равна $p_i q$, поэтому искомая вероятность равна:

$$P_A = 1 - (1 - p_1 q) \dots (1 - p_N q).$$

Выясним, при каком распределении вероятности между p_i будет достигнут максимум. Рассмотрим функцию $f(p) = (1 - pq)(1 - (C - p)q)$ на минимум. Он достигается в максимально удаленной от вершины этой параболы, $\frac{C}{2}$, точке. Поскольку $0 \le p, C - p \le 1$, то при $C \le 1$ имеем $0 \le p \le C$ и этот минимум достигается при $p \in \{0, C\}$, а при $1 < C \le 2$ имеем $C - 1 \le p \le 1$ и этот минимум достигается при $p \in \{1, C - 1\}$ Таким образом, максимум величины P_A достигается при $p_{i_1} = \dots = p_{i_k} = 1$, для некоторых $i_j \le N$, а остальные нули, и равен

$$1 - (1 - q)^k.$$

В случае Fork-атаки рассуждения проводятся аналогично с заменой p_i на $p_i\left(1 - \frac{a}{100}\right)$, что приводит к тем же результатам с заменой q на $q\left(1 - \frac{a}{100}\right)$. Таким образом, получена следующая теорема

Теорема 1. Выбором достаточно большого k можно избежать влияния любой Fork-атаки на параметр P_A. А именно, нужно выбрать скорректированное значение k^* так, чтобы $(1 - q)^k \ge \left(1 - \left(1 - \frac{a}{100}\right)q\right)^{k^*}$ или

$$k \ge k^* log_{1-q}\left(1 - \frac{a}{100}\right)(1 - q).$$

Б) Искомая вероятность, очевидно, равна $P_{B,j} = (1 - P_A)^{j-1} P_A$. Поэтому свойства этого параметра аналогичны предыдущему и следуют из теоремы 1.

В) Поскольку

$$\sum_{j=1}^{\infty} P_{Б,j} = P_A \sum_{j=1}^{\infty} (1-P_A)^{j-1} = P_A \frac{1}{1-(1-P_A)} = 1$$

то по определению имеем:

$$M = \sum_{j=1}^{\infty} j P_{Б,j} = P_A \sum_{j=1}^{\infty} j (1-P_A)^{j-1} =$$

$$= P_A \left(\frac{1}{1-(1-P_A)} + \frac{1-P_A}{1-(1-P_A)} + \cdots \right) =$$

$$1 + (1-P_A) + (1-P_A)^2 + \cdots =$$

$$= \frac{1}{1-(1-P_A)} = \frac{1}{P_A}.$$

А для дисперсии: D =

$$= P_A \sum_{j=1} j^2 (1-P_A)^{j-1} - \frac{1}{P_A^2} =$$

$$= \frac{2-P_A}{P_A^2} - \frac{1}{P_A^2} = \frac{1-P_A}{P_A^2}.$$

$$\sqrt{D} = \frac{\sqrt{1-P_A}}{P_A}$$

Отметим, что , то есть среднее отклонение меньше, чем матожидание.

Г) Рассматриваемая скорость равна единице минус среднее число отброшенных блоков за шаг, а именно

$$1 - [qp_1 1\,(1-qp_1 2) \ldots (1-qp_1 N) +$$

$$+ 2qp_2 (1-qp_3) \ldots (1-qp_N) + \cdots + Nqp_N].$$

Решим данную задачу при условии $p_i = p = \frac{N}{k}, i = 1, \ldots, N.$ Имеем

$$1 - qp\left((1-qp)^{N-1} + 2(1-qp)^{N-2} + \cdots + N\right) =$$

$$= 1 - qp(1-qp)^{N-1}(1 + 2(1-qp)^{-1} +$$

$$\ldots + N(1-qp)^{-(N-1)}).$$

Для вычисления этой суммы заметим, что если $f(a) = 1 + a + \cdots a^N$, то

$$f(a) = \frac{1-a^{N+1}}{1-a},$$

$$f'(a) = 1 + 2a + 3a^2 + \cdots + Na^{N-1} =$$

$$= \frac{-(N+1)a^N(1-a) + 1 - a^{N+1}}{(1-a)^2}.$$

Используя эту формулу при $a = \frac{1}{1-pq}$, получим следующее выражение для нашей суммы:

$$1 - qp(1-qp)^{N-1} \cdot$$

$$\cdot \frac{1 + \left(\frac{1}{1-qp}\right)^N \left(N \frac{1}{1-qp} - (N+1)\right)}{(1 - \frac{1}{1-qp})^2} =$$

$$= 1 - qp(1-qp)^{N+1} \cdot$$

$$\cdot \frac{1 + \frac{1}{(1-qp)^N}\left(N \frac{1}{1-qp} - (N+1)\right)}{(qp)^2} =$$

$$1 - \frac{(1-qp)^{N+1} + N - (N+1)(1-qp)}{qp} =$$

$$= 1 - \frac{(1-qp)^{N+1} - 1 + (N+1)qp}{qp} =$$

$$= \frac{1}{qp} - N - \frac{(1-qp)^{N+1}}{qp},$$

что при условии достижения максимума в теореме 1, $N = k, p = 1$, равно

$$\frac{1-(1-q)^{k+1}}{q} - k$$

Отсюда вытекает условие на q, при выполнении которого скорость развития дерева блокчейна больше нуля, а именно:

$$\frac{1-(1-q)^{k+1}}{q} > k.$$

При q, стремящемся к бесконечности, для левой части этого неравенства имеем

$$\frac{1-\left(1-(k+1)q + O(q^2)\right)}{q} = k + 1 + O(q).$$

Поэтому $q \in (0,1)$, удовлетворяющее указанному неравенству, существует, что косвенно подтверждает адекватность нашей модели.

В случае Fork-атаки аналогично получим $\left(1 - q\left(1 - \frac{a}{100}\right)\right)^{k'}$ вместо $(1-q)^k$, откуда при корректировке k как в Теореме 1, получим тот же результат.

Д) Пусть блок с номером -1 плохой. Вероятность того, что он не будет распознан, очевидно, равна $P_E = (1-p_1)(1-p_2)\ldots(1-p_N)$. Поскольку максимум функции $f(p) = (1-p)(1-(C-p))$

достигается в точке $\frac{C}{2}$, то, как и выше находим, что P_E достигает своего максимума при равных p_i, то есть при $p_i = \frac{k}{N}$ и будет в этом случае равна $\left(1-\frac{k}{N}\right)^N \approx e^{-k}$, а Fork-атака может быть сведена на нет, как и выше, выбором значений N^*, k^* так, чтобы $\left(1-\frac{k}{N}\right)^N = \left(1-\frac{\left(1-\frac{a}{100}\right)k'}{N'}\right)^{N'}$. То есть при больших N, N^* можно выбрать $k^* = \frac{100k}{100-a}$, или $k^* = Mk$ при $a = \frac{M-1}{M}$.

V. Заключение.

Приведенные оценки показывают, что наша формальная модель блокчейна устойчива к Fork-атаке в том смысле, что её влияние может быть сведено к нулю выбором параметров безопасности. Конечно, мы предполагаем, что ”честные” абоненты уведомлены в том, что такая атака началась и имеют возможность адекватно скорректировать свои параметры безопасности.

Получено условие на параметры безопасности, обеспечивающее положительность скорости развития дерева блокчейна. Отметим, что это является необходимым условием при практическом использовании блокчейна для формирования и хранения защищённых баз данных.

БИБЛИОГРАФИЯ

[1] S. Nakamoto: ”Bitcoin: A Peer-to-Peer Electronic Cash System” Available: www.bitcoin.org/bitcoin.pdf

[2] М.А. Черепнёв “Криптографические протоколы” Москва: МАКС Пресс, 2018.

[3] R. Pass, E. Shi “Hibrid Consensus: E-client Consensus in the Permissionless Model” Available: e-print arxiv 2016/917

Estimates of Fork-attack effectiveness on blockchain protocol

M.A. Cherepniov

***Abstract*--- We consider protocol of secure construction and storage of data base, named "blockchain". This protocol may be considered as some development of control sums and electronic signature schemes on the case when all participants of the set guarantees consistency of the data in the same manner. We construct a formal model of blockchain algorithm. We propose, that so called "long chain condition" is a necessary part of the considered blokchain protocol. We also propose, that when user finds an error, the corresponding block reject with all blocks after it in blockchain tree. For this model we obtained some probability estimates of the events, like error correcting of blocks. We obtain the estimate of mean value of speed of blockchain growth. We demonstrate that modification of these estimates, when some fixed quantity of "adversaries" provide bad blocks may be minimize by "honest" users. We show that "honest" users may protect against Fork-attack by choosing security parameters, like number of checked blocks in one step. To the other hand in this case, speed of growth of blockchain tree may become negative, and our protocol fall down. We present quantitative bounds.**

***Key words*---blockchain, Fork-attack, blockchain tree growth speed.**

REFERENCES

[1] S. Nakamoto: "Bitcoin: A Peer-to-Peer Electronic Cash System" Available: www.bitcoin.org/bitcoin.pdf
[2] M.A. Cherepnjov "Kriptograficheskie protokoly" Moskva: MAKS Press, 2018.
[3] R. Pass, E. Shi "Hibrid Consensus: E-client Consensus in the Permissionless Model" Available: e-print arxiv 2016/917