Юрай Громкович,
Аблаев Фарид Мансурович

АЛГОРИТМЫ И СЛУЧАЙНОСТЬ УРОК 3. АНАЛИЗ ПРОТОКОЛА WITNESS И ЕГО ОБОБЩЕНИЕ

В этой статье мы разберем упражнения из предыдущей статьи [1], проведем анализ протокола WITNESS и рассмотрим его обобщение протокол WITNESS(10). Начнем с анализа упражнений.

Упражнение 1. Нам необходимо сравнить две последовательности x и y с помощью протокола one-bit-WITNESS. По условию последовательности x и y длины $n = 10^{16}$ различны.

Рассмотрим случай, когда x и y различны в 10^5 битах.

Протокол one-bit-WITNESS сработает на x и y правильно, если Alice случайным образом выберет один из битов, в которых последовательности x и y различны. Следовательно вероятности $P(x \neq y)$ и $P(x = y)$ правильной и ошибочной работы протокола будут

$$P(x \neq y) = \frac{10^5}{10^{16}} = \frac{1}{10^{11}} = 10^{-11}, \qquad (1)$$

$$P(x = y) = 1 - 10^{-11} \qquad (2)$$

соответственно.

Теперь подсчитаем для протокола one-bit-WITNESS число передаваемых бит по формуле (1) из предыдущей статьи [1]:

$$\left\lceil \log_2 10^{16} \right\rceil + 1 = 16 \cdot 4 + 1 = 65. \qquad (3)$$

Упражнение 2. Проведем анализ работы протокола one-bit-WITNESS(10) для последовательностей x и y из предыдущего упражнения. По условию последовательности x и y различны в 10^5 битах и одинаковы в $(10^{16} - 10^5)$ битах.

Протокол one-bit-WITNESS(10) выдает ответ «**последовательности *x* и *y* не равны**», когда хотя бы один из десяти вариантов работы протокола one-bit-WITNESS заканчивался результатом «**последовательности *x* и *y* не равны**», и ответ «**последовательности *x* и *y* равны**», если все десять результатов работы протокола one-bit-WITNESS – «**последовательности *x* и *y* равны**».

Итак, вероятность $P'(x = y)$ получить ответ «**последовательности *x* и *y* равны**» при работе протокола one-bit-WITNESS(10) – это вероятность получения результата «**последовательности *x* и *y* равны**» в каждом из 10 независимых применений протокола one-bit-WITNESS. В силу (3) мы получаем, что вероятности $P'(x = y)$ и $P'(x \neq y)$ получить ответы «**последовательности *x* и *y* равны**» и «**последовательности *x* и *y* не равны**» при работе протокола one-bit-WITNESS(10) соответственно равны:

$$P'(x = y) = P(x = y)^{10} = (1 - 10^{-11})^{10}.$$

$$P'(x \neq y) = 1 - P(x = y)^{10} = 1 - (1 - 10^{-11})^{10}.$$

Теперь подсчитаем число передаваемых бит. Протокол one-bit-WITNESS(10) повторяет протокол one-bit-WITNESS 10 раз, поэтому мы будем передавать в десять раз больше бит, чем протокол one-bit-WITNESS. Поэтому, в силу соотношения (3), протокол one-bit-WITNESS(10) при обработке наших последовательностей x и y предает $10 \cdot 65 = 650$ бит.

Подведем итоги рассмотрений протокола one-bit-WITNESS, one-bit-WITNESS(10) и упражнения 2. Протоколы one-bit-WITNESS, one-bit-WITNESS(10) работают, как это принято говорить, с односторонней ошибкой: если анализируемые последовательности x, y равны, то протоколы не ошибаются, однако если $x \neq y$, то ошибки протоколов могут быть весьма велики.

Упражнение 2 показывает, что вероятность $P'(x = y)$ ошибки протокола one-bit-WITNESS(10) меньше, чем вероятность $P'(x = y)$ ошибки протокола one-bit-WITNESS. Это свойство является следствием повторения протокола one-bit-WITNESS. Оказывается, это общая тенденция, которая проявляется в самых разных областях информатики и техники – «резервирование увеличивает надежность».

Перейдем к анализу протокола WITNESS.

Упражнение 3. Напомним, что при помощи протокола WITNESS мы сравниваем последовательности $x = 001111$ и $y = 010110$.

Вопросы 1, 2. В множестве $PRIM(6^2)$ не существует числа, отличного от 7, выбор которого привел бы к неправильному ответу протокола WITNESS. Этот факт можно проверить путем простого перебора. Но есть и другой способ решения, отличный от перебора всех вариантов. Его мы применим ниже при анализе упражнения 5 и протокола WITNESS.

Вопрос 3. В множестве $PRIM(6^2)$ 11 чисел, из них только одно «плохое», которое мы можем выбрать с вероятностью 1/11. Тогда «хорошее» число, на котором мы получим правильный результат, можно выбрать с вероятностью 10/11.

Вопрос 4. *Alice* должна будет передать простое число p и остаток s от деления числа $Number(x)$ на это простое число. Покажем, что для передачи чисел p и s достаточно девяти бит.

Действительно, множество $PRIM(6^2)$ содержит 11 чисел, поэтому $\lceil \log_2 11 \rceil = 4$ бита достаточно для взаимнооднозначного двоичного кодирования чисел из $PRIM(6^2)$.

Далее, остаток s от деления на число p не может быть больше делителя. В множестве $PRIM(6^2)$ самым большим числом является число 31, значит, для взаимно однозначного двоичного кодирования чисел $s < 31$ достаточно $\lceil \log_2 31 \rceil = 5$ бит.

В итоге получаем, что *Alice* достаточно девяти бит для передачи сообщения.

Упражнение 4. Для двух одинаковых чисел остатки от деления на одно и то же число всегда буду равны, поэтому не существует простого числа, которое привело бы к неправильной работе протокола.

Упражнение 5. Напомним, что $x = 10011011$ и $y = 01010101$.

1. Мы интерпретируем x и y как двоичные последовательности, поэтому имеем

$$Number(x) = 2^7 + 2^4 + 2^3 + 2^1 + 2^0 = 155,$$
$$Number(y) = 2^6 + 2^4 + 2^2 + 2^0 = 85.$$

Множество $PRIM(8^2)$ состоит из всех простых чисел, не превышающих 8^2:

$$PRIM(8^2) = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 57, 59, 61\}.$$

Найдем все числа из $PRIM(8^2)$, которые приведут к неправильной работе протокола WITNESS. Их уже 19, перебор гарантирует верное решение, но он утомителен. Рассмотрим другой подход.

Протокол WITNESS ошибется, если для сравнения чисел $Number(x)$ и $Number(y)$ будет выбран модуль $m \in PRIM(8^2)$ такой, что остатки s и q от деления чисел $Number(x)$ и $Number(y)$ на m будут совпадать ($s = q = r$), то есть если числа $Number(x)$ и $Number(y)$ представимы в виде:

$$Number(x) = a_1 m + r,$$
$$Number(y) = a_2 m + r$$

для подходящих a_1 и a_2. Другими словами разность

$$Number(x) - Number(y) = 155 - 85 = 70$$

кратна выбранному простому числу m.

Известно, что любое целое число однозначно представимо в виде произведения степеней простых чисел. Этот факт носит название основной теоремы теории чисел. В частности, для числа 70 имеем $70 = 2 \cdot 5 \cdot 7$. Значит, числа 2, 5 и 7 и только они из множества $PRIM(8^2)$ приведут к неправильной работе протокола.

Вопрос 2. Множество $PRIM(8^2)$ состоит из 19 чисел, три из которых приводят к неправильному результату. Значит, при выборе остальных 16 чисел мы получим правильный результат. Следовательно, при равновероятностном выборе числа $m \in PRIM(8^2)$ вероятность правильной об-

работки последовательностей x и y протоколом WITNESS равна 16/19.

Вопрос 3. *Alice* должна передать номер выбранного простого числа и остаток от деления.

Для двоичного кодирования числа m из множества $PRIM(8^2)$, состоящего из 19 чисел, достаточно $\lceil \log_2 19 \rceil = 5$ бит. Для двоичного кодирования числа s, которое не превышает числа 61, достаточно $\lceil \log_2 61 \rceil = 6$ бит. Итак, Alice для передачи информации при работе в рамках протокола WITNESS достаточно 11 бит.

АНАЛИЗ КОММУНИКАЦИОННОГО ПРОТОКОЛА WITNESS

Компьютеры Alice и Bob содержат двоичные последовательности $x = x_1 \ldots x_{n-1} x_n$ и $y = y_1 \ldots y_{n-1} y_n$. В соответствии с протоколом WITNESS, Alice образует для себя вспомогательное множество $PRIM(n^2)$.

• *Вероятностный этап.* Компьютер Alice с равной вероятностью $\frac{1}{|PRIM(n^2)|}$ выбирает[1] число p из множества $PRIM(n^2)$.

• *Детерминированный этап.* Alice вычисляет число s – остаток от деления числа $Number(x)$ на число p. Компьютер Alice отправляет компьютеру Bob выбранное число p и вычисленное число s.

Компьютер Bob, получив числа p и s, производит следующие действия:

1) вычисляет остаток q от деления числа $Number(y)$ на число p.

2) если $s = q$, тогда Bob выдает ответ **«последовательности x и y равны»**; если $s \neq q$, тогда Bob выдает ответ **«последовательности x и y не равны»**.

Начнем с анализа числа передаваемых протоколом битов. Длина двоичного представления чисел $Number(x)$ и $Number(y)$ не больше чем n. Протокол WITNESS отправляет по каналу связи два целых числа: простое число p, не большее n^2, и остаток s от деления $Number(x)$ на p. Натуральное число, которое не больше чем n^2, можно представить $\lceil \log_2 n^2 \rceil \leq 2 \cdot \lceil \log_2 n \rceil$ битами.

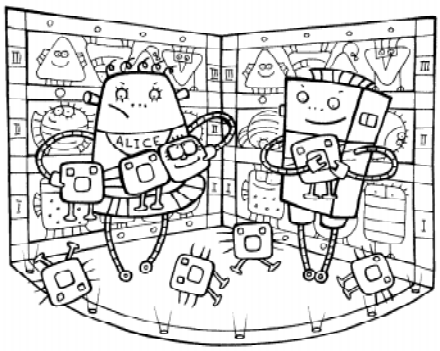
Тогда *Alice* может использовать ровно[2] $2\lceil \log_2 n \rceil$ битов для передачи каждого числа. В таком случае для любых входных данных длины n коммуникационная сложность будет равна $4 \cdot \lceil \log_2 n \rceil$.

При $n = 10^{16}$ WITNESS будет передавать по каналу связи

$$4 \cdot \lceil \log_2 10^{16} \rceil \leq 4 \cdot 16 \cdot \lceil \log_2 10 \rceil = 256$$

коммуникационных битов. Как было уже сказано ранее, сложность лучших детерминированных протоколов для этого случая будет не менее 10^{16}. Разница между передачей 256 и 10^{16} битов огромна. Даже если была бы возможность надежной передачи 10^{16} битов, выигрыш в стоимости передачи будет невероятным. И за это нам придется заплатить тем, что мы не всегда будем получать верный результат.

Теперь проанализируем, с какой вероятностью этот протокол может выдать неправильный результат для произвольных входных данных. Рассмотрим два случая. Первый случай – последовательности x и y равны, второй – последовательности различны.



Алис вычисляет число s – остаток от деления Number(x) на p... отправляет компьютеру Bob ... вычисленное число s

[1] Через $|PRIM(n^2)|$ будем обозначать число элементов (мощность) множества $PRIM(n^2)$.

[2] Если двоичное представление будет меньше чем $2\lceil \log_2 n \rceil$ битов, то можно добавить в начало несколько нулей и тем самым достичь нужной длины.

Случай 1. Если последовательности x и y одинаковы, то протокол WITNESS всегда выдаст правильный ответ, так как мы интерпретируем последовательности x и y как числа и для любых двух одинаковых чисел остаток от деления на одно и то же число всегда будет одинаковым. Значит, в первом случае вероятность выдать правильный ответ **«последовательности x и y равны»** – единица, а вероятность выдать неправильный ответ равна нулю.

Случай 2. $x \neq y$. Здесь мы обобщим подход, примененный в упражнении 5. Пусть числа $p_1, ..., p_t$ – это все числа из множества $PRIM(n^2)$, для которых выполняется:

$$\begin{aligned} &Number(x) = Number(y) \pmod{p_1} \\ &Number(x) = Number(y) \pmod{p_2} \\ &\quad\vdots \\ &Number(x) = Number(y) \pmod{p_t} \end{aligned} \qquad (4)$$

При выборе этих чисел $p_1, p_2, ..., p_t$ протокол WITNESS ошибется при сравнении последовательностей x, y. И вероятность $P_{err}(x, y)$ ошибки протокола WITNESS на последовательностях x, y равна

$$P_{err}(x, y) = \frac{t}{|PRIM(n^2)|}. \qquad (5)$$

Оценим величину $P_{err}(x, y)$.

Оценка знаменателя. В силу теоремы о распределении простых чисел (см. [2]), применяя формулу Чебышева

$$Prim(m) > 0.9\frac{m}{\ln m}$$

получаем:

$$Prim(n^2) > 0.45\frac{n^2}{\ln n}.$$

Оценка числителя. Покажем, что $t \le n$.

Положим $K = |Number(x) - Number(y)|$. С одной стороны, в силу системы равенств (4), каждое из простых чисел $p_1, p_2, ..., p_t$ делит число K. Поэтому имеем:

$$p_1 \cdot p_2 \cdot ... \cdot p_t \le K.$$

С другой стороны, в силу того, что $Number(x) \le 2^n$ и $Number(y) \le 2^n$, имеем:

$$K \le 2^n.$$

Объединяя эти два неравенства, получаем. что $p_1 \cdot p_2 \cdot ... \cdot p_t \le 2^n$.

А это означает, что $t \le n$.

Итак, вероятность $P_{err}(x, y)$ ошибки протокола WITNESS при сравнении x, y, учитывая полученные выше оценки величин $PRIM(n^2)$ и t, оценивается сверху следующим образом:

$$P_{err}(x, y) = \frac{t}{|PRIM(n^2)|} \le \frac{\ln n}{0.45n} < \frac{2\ln n}{n}.$$

Полученная оценка вероятности ошибки $P_{err}(x, y)$ показывает, что при увеличении длины n сравниваемых последовательностей x и y, вероятность ошибочной работы протокола WITNESS уменьшается.

WITNESS(10)

Компьютер Alice содержит n битов $x = x_1 ... x_n$, и компьютер Bob содержит n битов $y = y_1 ... y_n$. Протокол WITNESS(10) сравнивает эти последовательности следующим образом.

• *Вероятностный этап.* Alice случайным образом выбирает 10 простых чисел $p_1, p_2, ..., p_{10}$ из $PRIM(n^2)$.

• *Детерминированный этап.* Для каждого числа p_i из $\{p_1, ..., p_{10}\}$ Alice вычисляет число

$$s_i = Number(x) \bmod p_i$$

и отправляет Bob двоичное представление двадцати чисел

$$p_1, p_2, ..., p_{10}, s_1, s_2, ..., s_{10}.$$

Получив эти двадцать, чисел компьютер Bob вычисляет десять чисел

$$q_i = Number(y) \bmod p_i$$

для $i = 1, 2, ...,10$. Если существует хотя бы одна такая пара s_i, q_i, что $q_i \neq s_i$, то Bob выдает ответ **«последовательности x и y не равны»**.

Если $q_i = s_i$ для всех i из 1, 2, ...,10, то Bob выдает ответ **«последовательности x и y равны»**.

АНАЛИЗ КОММУНИКАЦИОННОГО ПРОТОКОЛА WITNESS(10)

Если $x = y$, то WITNESS(10) выдает правильный результат **«последовательности x и y равны»** точно так же, как и протокол

WITNESS. Если $x \neq y$, тогда WITNESS(10) может выдать неправильный результат, только если ни одно из 10 выбранных простых чисел $p_1, p_2,..., p_{10}$ не является свидетелем неравенства чисел $Number(x)$ и $Number(y)$. Достаточно, чтобы по крайней мере одно из 10 простых чисел было таким свидетелем, например p_4, тогда Bob получит, что $s_4 \neq q_4$, а значит, $x \neq y$. Так как вероятность выбрать несвидетеля неравенства чисел $Number(x)$ и $Number(y)$ есть вероятность $P_{err}(x, y)$ ошибки протокола WITNESS, которая. как мы выяснили выше не превышает величины $2\ln n/n$, то вероятность выбрать при десяти независимых испытаниях десять несвидетелей неравенства чисел $Number(x)$ и $Number(y)$ не больше чем

$$\left(\frac{2\ln n}{n}\right)^{10} = \frac{2^{10}\cdot(\ln n)^{10}}{n^{10}}.$$

Для $n = 10^{16}$ эта вероятность не больше

$$\frac{0.4714}{10^{141}}.$$

Насколько мала эта вероятность?

Произведение возраста Вселенной в секундах на количество протонов во Вселенной меньше чем 10^{99}.

Таким образом, мы уменьшили вероятность ошибки до уровня ниже любого разумного предела, а значит, выполнили все требования, требуемые на практике. И за столь невероятный выигрыш мы платим очень маленькую цену. Вычисления для 10 простых чисел вместо одного увеличивают расходы на связь на порядок. Коммуникационная сложность протокола WITNESS(10) равна $40\cdot\lceil\log 2n\rceil$.

Этими расходами на связь можно пренебречь. Например, для $n = 10^{16}$ WITNESS(10) передает только 2560 бит.

Поскольку протокол WITNESS(10) можно рассматривать как 10 представлений WITNESS, то можно сказать, что *при увеличении количества попыток сложность растет линейно вместе с количеством повторений (попыток найти свидетеля), в то время как вероятность ошибки уменьшается с экспоненциальной скоростью.*

По сути, эта ситуация относится к самым благоприятным из тех, когда идет поиск алгоритма для решения задачи.

Наш рандомизированный протокол WITNESS(10) очень хорош по двум причинам. Во-первых, уже протокол WITNESS обеспечивал высокую степень надежности. Во-вторых, метод «повторения эксперимента» (повторного случайного поиска свидетеля) существенно снижает вероятность ошибки даже в том случае, когда вероятность выбрать свидетеля мала.

Литература

1. *Громкович Ю., Аблаев Ф.М.* Алгоритмы и случайность. Урок 2. Когда случайность может быть полезной // Компьютерные инструменты в школе, 2012. № 2. С. 30–35.

2. Теорема о распределении простых чисел – Википедия http://ru.wikipedia.org/wiki/.

Юрай Громкович,
Professor of Computer Science, Swiss Federal Institute of Technology, Zürich,

Аблаев Фарид Мансурович,
доктор физико-математических наук, профессор, заведующий кафедрой теоретической киберенетики Казанского федерального университета.