Доказательства с нулевым знанием

Алиса знает решение некоторой сложной задачи, она хочет убедить Боба в этом, однако так, чтобы Боб не узнал самого решения задачи.

Боб должен убедиться в том, что Алиса знает решение, но не должен узнать что-нибудь о самом решении.

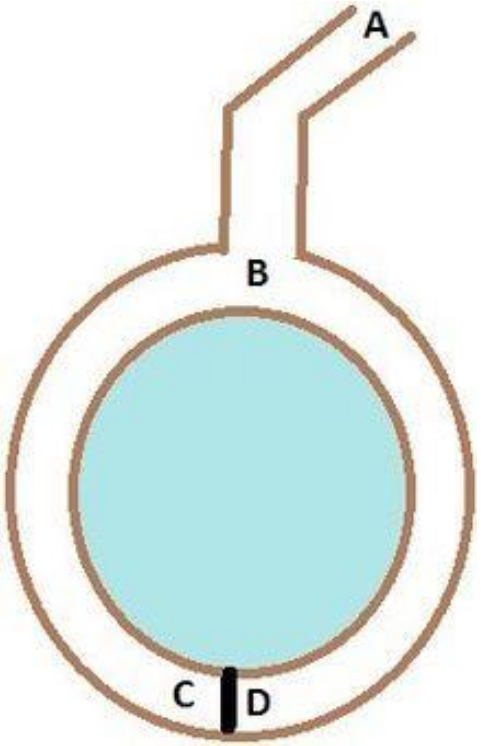
- Задача актуальна для компьютерных сетей: Боб (сервер или контроллер домена) должен принять решение о допуске Алисы к информации, хранящейся в сети, но при этом Алиса не хочет, чтобы кто-либо, прослушивающий канал передачи данных и сам сервер, получил какие-либо знания о ее пароле.

- Наша задача — построить протокол «доказательства с нулевым знанием». При этом считаем, что каждый из участников может вести «нечестную» игру и пытаться обмануть другого.

В качестве сложных задач рассмотрим NP-полные. NP-полнота задачи неформально означает, что время решения задачи растет экспоненциально с ростом размера задачи (объема исходных данных).

- Пещера нулевого разглашения
- Задача раскраски графа тремя красками
- Задача нахождения гамильтонова цикла в графе
- Протокол Фиата — Шамира

Пещера нулевого разглашения



P (Peggy) — от англ. *prover* (доказывающий)

V (Victor) — от англ. *verifier* (проверяющий)

- Пегги знает «ключ», открывающий дверь между C и D.

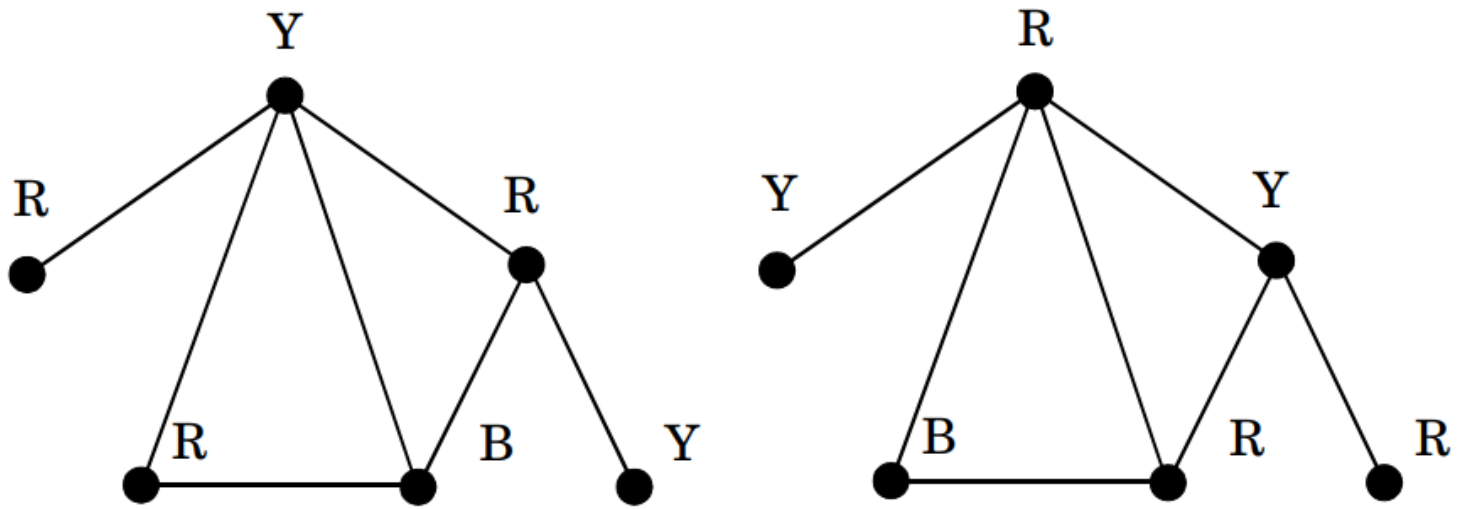
Виктор хочет узнать, действительно ли Пегги знает пароль,

при этом Пегги не хочет выдавать сам пароль.

- Пещера имеет круглую форму.
- Пока Виктор находится в точке A, Пегги идёт к двери. После того, как она исчезает из виду, Виктор идёт в точку B, и кричит оттуда: «Пегги нужно выйти справа» или «Пегги нужно выйти слева».
- Каждый раз вероятность того, что Пегги не знает пароль, равна 50 %.

Задача о раскраске графа

В задаче о раскраске графа рассматривается граф с множеством вершин V и множеством ребер E (числа элементов в этих множествах будем обозначать через $|V|$ и $|E|$). Алиса знает правильную раскраску этого графа тремя красками (красной (R), синей (B) и желтой (Y)). Правильная раскраска — это такая, когда любые две вершины, соединенные одним ребром, окрашены разными цветами.



а *б*

Рис. 5.1. Примеры раскрасок: *а* — правильная, *б* — неправильная

Для получения правильной раскраски графа тремя красками известны только экспоненциальные алгоритмы, т.е. такие, у которых время решения растет экспоненциально с ростом числа вершин и ребер в графе. Поэтому в случае больших $|V|$ и $|E|$ эта задача практически неразрешима.

Итак, Алиса знает (правильную) раскраску графа с большими $|V|$ и $|E|$. Она хочет доказать это Бобу, но так, чтобы он ничего не узнал об этой раскраске.

Протокол доказательства состоит из множества одинаковых этапов. Опишем сначала один этап.

Шаг 1. Алиса выбирает случайно перестановку Π из трех букв R, B, Y и перенумеровывает все вершины графа согласно этой перестановке. Очевидно, что раскраска останется верной. Например, если $\Pi = (\mathrm{Y}, \mathrm{R}, \mathrm{B})$, то граф слева на рис. 5.1 превращается в граф на рис. 5.2.

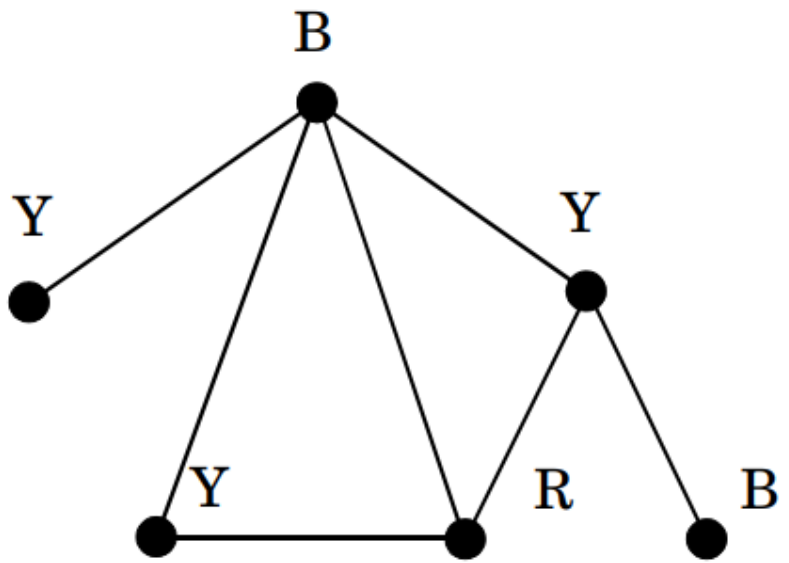


Рис. 5.2. Другой вариант раскраски

Шаг 2. Для каждой вершины v из множества V Алиса генерирует большое случайное число r и заменяет в нем два последних бита на 00, что соответствует красной вершине, 01 — синей, 10 — желтой.

Шаг 3. Для каждой вершины v Алиса формирует данные, используемые в RSA, а именно, P_v, Q_v, $N_v = P_v Q_v$, c_v и d_v.

Шаг 4. Алиса вычисляет

$$Z_v = r_v^{d_v} \bmod N_v$$

и посылает Бобу значения N_v, d_v и Z_v для каждой вершины графа.

Шаг 5. Боб выбирает случайно одно ребро из множества E и сообщает Алисе, какое именно ребро он выбрал . В ответ Алиса высылает числа c_{v_1} и c_{v_2}, соответствующие вершинам этого ребра. После этого Боб вычисляет

$$\hat{Z}_{v_1} = Z_{v_1}^{c_{v_1}} \bmod N_{v_1} = r_{v_1}, \quad \hat{Z}_{v_2} = Z_{v_2}^{c_{v_2}} \bmod N_{v_2} = r_{v_2}$$

и сравнивает два младших бита в полученных числах. При правильной раскраске два младших бита в числах $\hat{Z}_{v_1}$ и $\hat{Z}_{v_2}$ должны быть различны. Если значения совпали, значит, Алиса пыталась обмануть Боба, и на этом все заканчивается. Если не совпали, то весь описанный процесс повторяется $a|E|$ раз, где $a > 0$ — параметр.

Утверждение 5.2. *Если Алиса не располагает правильной раскраской графа, то вероятность того, что она сможет обмануть Боба не превышает* e^{-a}, *где* $e \approx 2.718$ — *число Эйлера (основание натурального логарифма).*

Замечание. Если взять большое a, то вероятность обмана можно сделать сколь угодно малой. Например, при $a = 5$ эта вероятность меньше 0.01.

Доказательство. Пусть Алиса не располагает правильной раскраской графа. Значит, хотя бы для одного ребра из E вершины окрашены в один цвет. Если Алиса будет действовать по протоколу, то вероятность того, что Боб обратится к такому ребру, не меньше $1/|E|$ (в этом случае Алиса разоблачена). Значит, вероятность того, что Алиса не разоблачена во время одного этапа, не превышает $1 - 1/|E|$ и, следовательно, вероятность того, что она не будет разоблачена за $a|E|$ этапов, не превышает $(1 - 1/|E|)^{a|E|}$. Используя известное неравенство $1 - x \leq e^{-x}$, получаем

$$(1 - 1/|E|)^{a|E|} \leq \left(e^{-1/|E|}\right)^{a|E|} = e^{-a}.$$

Проверим все свойства, необходимые для протокола с нулевым знанием.

1. Мы видим, что вероятность возможности обмана для Алисы может быть сделана сколь угодно малой.

2. Посмотрим, почему Боб не получает никакой информации о раскраске. Из-за того, что цвета переставляются случайно на каждом этапе (см. шаг 1), он не сможет узнать истинную раскраску, перебирая все ребра одно за другим, и вообще он ничего не узнает о правильной раскраске. То, что на втором шаге выбирается случайное число r_v, не позволяет Бобу вычислить по имеющимся N_v и d_v коды соответствующих красок. Он не может декодировать полученное Z_v потому, что он не знает чисел c_v, так как они для всех вершин не высылаются, а вычислить их он не может , не зная P_v и Q_v.

3. Рассмотрим еще одну возможность обмана, которая в принципе может быть у Алисы. Казалось бы, Алиса может подменить c_{v_1} и c_{v_2}, если ей это выгодно. Однако это невозможно в силу того, что число c_v, удовлетворяющее равенству

$$c_v d_v \bmod ((P_v - 1)(Q_v - 1)) = 1,$$

единственно.

Таким образом, выполнены все свойства:

1) Алиса доказывает Бобу, что знает решение задачи, и вероятность того, что Боб обманут, не больше e^{-a};

2) Боб не получает никаких сведений о раскраске.

Рассмотрим последнюю возможность обмана для всех участников. Что будет, если они будут уклоняться от указанного алгоритма, выбирая параметры не случайно?

Пусть, например, Боб запрашивает ребра графа не случайно, а по какому-нибудь простому правилу (например, в соответствии с их номерами). В этом случае, если у Алисы нет правильной раскраски, то она сможет обмануть Боба, «правильно» раскрашивая те ребра, которые будут запрошены. Таким образом, Боб заинтересован в том, чтобы его запросы были случайны и не содержали в себе какой-либо закономерности.

Стойкость остальных шагов определяется стойкостью RSA, и при больших P_v и Q_v система достаточно надежена.

Задача о нахождении гамильтонова цикла в графе

Любое математическое утверждение может быть представлено в виде графа, причем доказательство этого утверждения соответствует гамильтонову циклу в этом графе.

Поэтому наличие протокола доказательства с нулевым знанием для гамильтонова цикла означает, что доказательство любого математического утверждения может быть представлено в форме доказательства с нулевым знанием.

Определение 5.1. *Гамильтоновым циклом* в графе называется непрерывный путь, проходящий через все вершины графа ровно по одному разу.

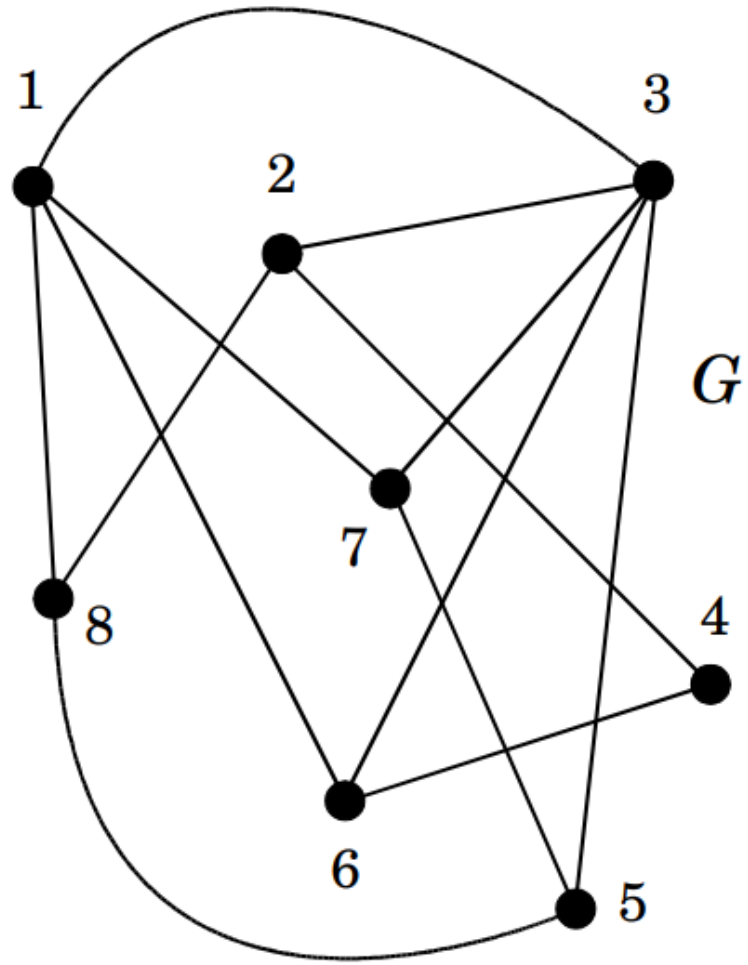


Рис. 5.3. Граф с гамильтоновым циклом (8, 2, 4, 6, 3, 5, 7, 1)

Путь, проходящий последовательно через вершины 8, 2, 4, 6, 3, 5, 7, 1, представляет собой гамильтонов цикл. Действительно, в этом пути содержатся все вершины графа, и каждая вершина посещается только один раз.

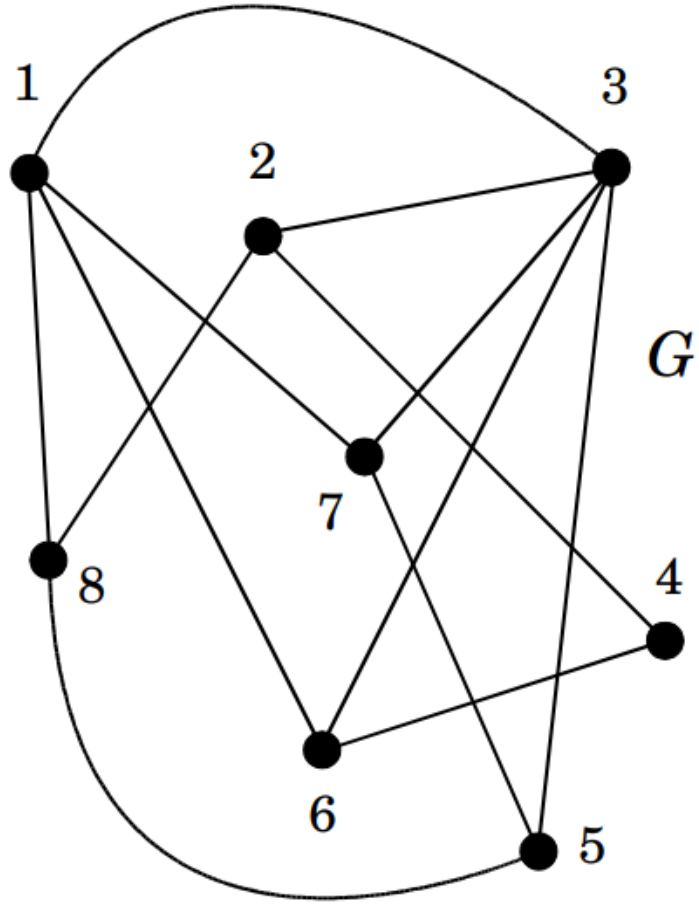


Рис. 5.3. Граф с гамильтоновым циклом (8, 2, 4, 6, 3, 5, 7, 1)

Ясно, что если в графе G с n вершинами гамильтонов цикл существует, то при некоторой нумерации вершин он пройдет точно через вершины с последовательными номерами 1, 2, 3, . . . , n.

Поэтому путем перебора всех возможных нумераций вершин мы обязательно найдем гамильтонов цикл. Но количество возможных нумераций равно n!, и поэтому уже при умеренно больших n, например, при n = 100, такой подход становится практически нереализуемым.

Доказано, что задача нахождения гамильтонова цикла в графе является **NP-полной**.

Неформально, NP-полнота рассматриваемой задачи означает, что для ее решения неизвестны алгоритмы существенно более быстрые, чем указанный метод перебора.

Наша задача: построение криптографического протокола, с помощью которого Алиса будет доказывать Бобу, что она знает гамильтонов цикл в некотором графе G так, чтобы Боб не получил никаких знаний о самом этом цикле.

«Нулевое знание»: независимо от числа реализаций протокола доказательства Боб будет располагать точно такими же сведениями о гамильтоновом цикле, какие он мог бы получить, просто изучая представленный ему граф G.

Допустим, что Алиса знает гамильтонов цикл в графе G. Теперь она может это доказывать Бобу и всем, кто имеет граф G, с помощью описываемого ниже протокола. Алиса может использовать это доказательство, например, для идентификации своей личности. Но прежде чем мы перейдем к описанию протокола, договоримся о некоторых обозначениях.

Мы будем обозначать графы буквами ***G***, ***H***, ***F***, понимая под этим одновременно соответствующие матрицы смежности.

Элемент матрицы $H_{ij} = 1$, если в графе ***H*** есть ребро, соединяющее вершины ***i*** и ***j*** ; $H_{ij} = 0$ в противном случае.

Символом || будем обозначать конкатенацию двух чисел, точнее, двоичных слов, им соответствующих. Нам понадобится шифр с **открытым ключом**.

Это может быть любой шифр, но для определенности будем использовать шифр **RSA**.

Будем считать, что Алиса сформировала систему RSA с открытыми параметрами N и d.

Важно, что зашифрованные в этой системе сообщения может расшифровать только Алиса и больше никто.

Протокол доказательства состоит из следующих четырех шагов (пояснения будут даны ниже).

Шаг 1. Алиса строит граф H, являющийся копией исходного графа G, где у всех вершин новые, случайно выбранные номера. На языке теории графов говорят, что H изоморфен G. Иными словами, H получается путем некоторой перестановки вершин в графе G (с сохранением связей между вершинами). Алиса кодирует матрицу H, приписывая к первоначально содержащимся в ней нулям и единицам случайные числа r_{ij} по схеме $\tilde{H}_{ij} = r_{ij} \| H_{ij}$. Затем она шифрует элементы матрицы $\tilde{H}$, получая зашифрованную матрицу F, $F_{ij} = \tilde{H}_{ij}^d \bmod N$. Матрицу F Алиса передает Бобу.

Шаг 2. Боб, получив зашифрованный граф F, задает Алисе один из двух вопросов.

1. Каков гамильтонов цикл для графа H?
2. Действительно ли граф H изоморфен G?

Шаг 3. Алиса отвечает на соответствующий вопрос Боба.

1. Она расшифровывает в F ребра, образующие гамильтонов цикл.
2. Она расшифровывает F полностью (фактически передает Бобу граф $\tilde{H}$) и предъявляет перестановки, с помощью которых граф H был получен из графа G.

Шаг 4. Получив ответ, Боб проверяет правильность расшифровки путем повторного шифрования и сравнения с F и убеждается либо в том, что показанные ребра действительно образуют гамильтонов цикл, либо в том, что предъявленные перестановки действительно переводят граф G в граф H.

Весь протокол повторяется t раз.

1. Зачем Алиса строит изоморфный граф? Если бы она этого не делала, то Боб, получив ответ на свой вопрос номер один, узнал бы гамильтонов цикл в графе G.

2. Зачем Алиса кодирует матрицу H? С этим приемом мы уже встречались при шифровании цветов вершин графа. Дело в том, что невозможно зашифровать непосредственно нули и единицы (с помощью шифра RSA они вообще не шифруются). Даже если заменить их на какие-то произвольные числа a и b, то мы получим всего два различных шифротекста, и Бобу не составит труда понять, какой из них какому числу соответствует. Т.е. структура графа не будет скрыта. Здесь мы сталкиваемся с типичной ситуацией, когда требуется использовать так называемый рандомизированный шифр. И такой шифр строится путем добавления случайных чисел в матрицу H перед шифрованием. Закодированная матрица $\tilde{H}$ точно также задает граф (нечетность числа означает наличие ребра, четность — его отсутствие), но после шифрования $\tilde{H}$ структура графа полностью скрывается (мы используем известное свойство шифра RSA — он полностью скрывает четность числа).

3. Зачем Боб задает два вопроса? Если бы он задавал только вопрос номер один, который по смыслу протокола является основным, то Алиса, не зная в действительности гамильтонова цикла в графе G, могла бы предъявить Бобу совсем другой граф с таким же количеством вершин и искуственно заложенным в него гамильтоновым циклом. Поэтому Боб иногда просит Алису доказать изоморфизм графов H и G. Важно, что Алиса не знает заранее, какой из двух вопросов задаст Боб.

4. Почему Боб не может задать сразу двух вопросов? В этом случае он узнал бы гамильтонов цикл в G, так как ему был бы показан гамильтонов цикл в H и правило перехода от H к G.

5. Зачем Боб проверяет правильность расшифровки? Если бы он этого не делал, то Алиса на четвертом шаге могла бы предоставить ему «выгодную» для себя информацию, а не ту, которую она посылала ему на втором шаге.

Более точно основные детали протокола обосновываются в ходе доказательства двух основных утверждений.

Утверждение 5.3. *Вероятность обмана при t реализациях протокола не превосходит 2^{-t}.*

Доказательство. Вначале покажем, что вероятность обмана в одной реализации протокола равна **1/2**. Заметим, что если Алиса действительно знает гамильтонов цикл в графе G, то она может правильно ответить на любой вопрос Боба. Если же она не знает гамильтонов цикл, то самое большее, что она может сделать, — это подготовиться к ответу на первый либо на второй вопрос. В ожидании первого вопроса, она создает новый граф с искусственно заложенным в него гамильтоновым циклом. Но в этом случае она не сможет доказать его изоморфизм графу G. В ожидании второго вопроса, она строит граф, изоморфный графу G. Но в этом случае она не сможет показать в нем гамильтонов цикл. Таким образом, вероятность успешности обмана равна вероятности угадывания номера вопроса. В предположении, что Боб задает оба вопроса с одинаковыми вероятностями, мы получаем, что вероятность обмана равна **1/2**.

Так как Боб прекращает игру при первом же неправильном ответе, вероятность обмана при t реализациях протокола не превосходит $(1/2)^t$.

Утверждение 5.4. *Представленный протокол реализует доказательство с нулевым знанием.*

Доказательство. Чтобы доказать, что Боб не получает никаких знаний в ходе реализации протокола, достаточно показать, что все, что он получает от Алисы, он мог бы получить сам, не вступая с ней ни в какое общение.

Рассмотрим вначале второй вопрос Боба. В ответ на этот вопрос он получает граф, изоморфный графу G. Но он сам мог строить сколько угодно изоморфных графов, и то, что присылает ему Алиса, это просто один из них.

Случай с первым вопросом не столь очевиден. В ответ на первый вопрос Боб получает гамильтонов цикл в графе, изоморфном графу G. На первый взгляд может показаться, что это дает Бобу какую-то информацию. Однако это не так.

Заметим, что если в G есть гамильтонов цикл, то при некоторой нумерации вершин существует изоморфный граф, который задается матрицей смежности вида:

$$\begin{pmatrix} * & 1 & * & \cdots & * & * & * \\ * & * & 1 & \cdots & * & * & * \\ & & & \cdots & & & \\ * & * & * & \cdots & * & 1 & * \\ * & * & * & \cdots & * & * & 1 \\ 1 & * & * & \cdots & * & * & * \end{pmatrix}, \qquad (5.8)$$

где $*$ означает неопределенность в наличии или отсутствии ребра.

Т.е. при такой нумерации гамильтонов цикл проходит через вершины в порядке возрастания номеров. Изменяя нумерацию вершин, Боб может получать из (5.8) всевозможные изоморфные матрицы. Когда Алиса, отвечая на его первый вопрос, открывает гамильтонов цикл, Боб видит как раз одну из таких матриц. Боб не получает от Алисы никакой информации, которую он не мог бы получить сам.

Пример 5.3. Возьмем в качестве основного граф G, изображенный на рис. 5.3. Его матрица смежности имеет вид

$$G = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & \boxed{1} \\ 0 & 0 & 1 & \boxed{1} & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & \boxed{1} & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \boxed{1} & 1 \\ 1 & 0 & \boxed{1} & 1 & 0 & 0 & 0 & 0 \\ \boxed{1} & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & \boxed{1} & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

В матрице с помощью $\boxed{\cdot}$ показан гамильтонов цикл.

Алиса выбирает некоторую случайную нумерацию вершин, например, 7, 4, 5, 3, 1, 2, 8, 6, и получает изоморфный граф

$$H = \begin{pmatrix} & 7 & 4 & 5 & 3 & 1 & 2 & 8 & 6 \\ 7 & 0 & 0 & 1 & 1 & \boxed{1} & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \boxed{1} \\ 5 & \boxed{1} & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 3 & 1 & 0 & \boxed{1} & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & \boxed{1} & 1 \\ 2 & 0 & \boxed{1} & 0 & 1 & 0 & 0 & 1 & 0 \\ 8 & 0 & 0 & 1 & 0 & 1 & \boxed{1} & 0 & 0 \\ 6 & 0 & 1 & 0 & \boxed{1} & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Для шифрования матрицы будем использовать систему RSA с параметрами $N = 55,\ d = 3$. Вначале закодируем матрицу H. В рамках данного примера просто припишем слева к каждому элементу матрицы выбираемую случайно с равными вероятностями цифру из множества $\{1, 2, 3, 4, 5\}$:

$$\tilde{H} = \begin{pmatrix} 50 & 20 & 11 & 31 & 21 & 40 & 20 & 10 \\ 40 & 30 & 50 & 20 & 10 & 41 & 50 & 21 \\ 41 & 30 & 50 & 11 & 30 & 20 & 51 & 40 \\ 11 & 10 & 41 & 30 & 51 & 41 & 30 & 21 \\ 31 & 20 & 40 & 11 & 50 & 10 & 41 & 31 \\ 50 & 41 & 20 & 21 & 40 & 10 & 21 & 50 \\ 40 & 30 & 31 & 50 & 41 & 21 & 30 & 40 \\ 20 & 41 & 10 & 51 & 41 & 20 & 30 & 40 \end{pmatrix}.$$

Теперь мы шифруем матрицу $\tilde{H}$, возводя каждый ее элемент в куб по модулю 55:

$$F = \begin{pmatrix} 40 & 25 & 11 & 36 & 21 & 35 & 25 & 10 \\ 35 & 50 & 40 & 25 & 10 & 06 & 40 & 21 \\ 06 & 50 & 40 & 11 & 50 & 25 & 46 & 35 \\ 11 & 10 & 06 & 50 & 46 & 06 & 50 & 21 \\ 36 & 25 & 35 & 11 & 40 & 10 & 06 & 36 \\ 40 & 06 & 25 & 21 & 35 & 10 & 21 & 40 \\ 35 & 50 & 36 & 40 & 06 & 21 & 50 & 35 \\ 25 & 06 & 10 & 46 & 06 & 25 & 50 & 35 \end{pmatrix}.$$

(При внимательном просмотре матрицы F может показаться, что использованный нами шифр плохо скрывает исходную матрицу H.

Это объясняется тем, что, во-первых, модуль 55 слишком мал и, во-вторых, в матрице $\tilde{H}$ много чисел, не взаимно простых с модулем. Для реальных систем RSA, где N — большое число, такая ситуация практически исключена.)

Боб получает матрицу F и задает один из двух вопросов. Если он просит доказать изоморфизм графов, то Алиса просто посылает ему кодированную матрицу $\tilde{H}$ и использованную нумерацию 7, 4, 5, 3, 1, 2, 8, 6. Боб проверяет соответствие матрицы $\tilde{H}$ матрице F, т.е.выполнение равенств $50^3 \bmod 55 = 40$, $20^3 \bmod 55 = 25$ и т.д. Из матрицы $\tilde{H}$ Боб получает граф H (просто отбросив старшую десятичную цифру). Затем он переставляет вершины графа G в соответствии с полученной нумерацией, как это делала Алиса, и убеждается в том, что H и G — один и тот же граф.

Если Боб просит показать ему гамильтонов цикл, то Алиса посылает ему соответствующий список (закодированных) ребер графа H: $(1, 5, 21), (5, 7, 41), (7, 6, 21), \ldots, (3, 1, 41)$. Каждый элемент содержит номера вершин и код ребра. Боб проверяет соответствие указанных в списке ребер матрице F, например, $21^3 \bmod 55 = 21 = F_{1,5}$, $41^3 \bmod 55 = 06 = F_{5,7}$ и т.д. Затем он убеждается, что указанный в списке путь проходит через все вершины графа по одному разу.