Блокчейн и распределенные реестры как виды баз данных

К. Д. Шилов,
научный сотрудник,
shilov-kd@ranepa.ru

А. В. Зубарев,
к. э. н., старший научный сотрудник, лаборатория математического моделирования экономических процессов
zubarev@ranepa.ru

Институт прикладных экономических исследований (ИПЭИ), Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (РАНХИГС)

Последние несколько лет технология распределенного реестра (distributed ledger technology, DLT), в частности блокчейн (blockchain), не только активно обсуждается, но и пилотно внедряется в большинстве развитых стран во многих областях экономики. Многие организации ставят своей целью разобраться в возможном применении технологии как для поиска путей развития бизнеса, так и для оптимизации текущих процессов для сокращения издержек. Бытует мнение, что технология распределенного реестра имеет значение, сравнимое с такими инновациями своего времени, как интернет и телеграф. Потенциал технологии будет полностью раскрыт в ближайшие 5-10 лет, однако внедряемые сегодня пилотные блокчейн-проекты в области финансов позволяют говорить об открывающихся новых бизнес-возможностях и кардинальной оптимизации существующих бизнес-процессов. Это, в том числе, свидетельствует в пользу гипотезы о глобальном изменении роли институтов и их места в экономике на горизонте нескольких десятилетий из-за массового внедрения технологии. Данная статья раскрывает основные положения и некоторые технические особенности, лежащие в основе технологии распределенного реестра, необходимые для понимания возможности ее применения в различных областях экономики.

Ключевые слова: блокчейн, технология распределенного реестра, базы данных, криптовалюты.

Введение

В последние несколько лет дискуссия о криптовалютах стала одной из самых популярных. Игроки на финансовых рынках пытаются совершить выгодные вложения в различные криптовалютные инструменты, экономисты-теоретики обсуждают, насколько данный феномен укладывается в концепцию частных денег Фридриха Хайека, юристы обсуждают правовые аспекты первичного размещения криптоактивов (так называемые ICO, Initial Coin Offering). Основной же интерес широких масс до сих пор вызывает самая первая криптовалюта Bitcoin, созданная и описанная неким Сатоси Накамото в статье 2008 г. [1], капитализация которой на начало мая 2018 г. составляла порядка $156 млрд. В этой же статье Накамото впервые была формализована идея технологии, которая в последующем будет обозначена термином блокчейн (blockchain — цепочка блоков) и будет использована в качестве технологической основы Bitcoin, а также многих других криптовалют.

Так что же собой представляет технология распределенного реестра? По сути, это лишь некоторый особый вид базы данных. Для более подробного раскрытия этого тезиса в первой части статьи мы рассматриваем традиционные виды базы данных и то, как распределенный реестр соотносится с ними. Во второй части статьи объясняется механизм работы блокчейна на примере криптовалюты биткоин. В третьей части приведена классификация видов распределенного реестра и обзор альтернативных механизмов консенсуса.

1. Традиционные базы данных и технология распределенного реестра

База данных, по определению, — это организованный, систематизированный набор некоторой информации. Самым распространенным на сегодняшний день видом баз данных является реляционная база данных (relational database). Обычную базу данных можно представить в виде некоторой таблицы (сущности), состоящую из столбцов и строк (кортежей). Реля-

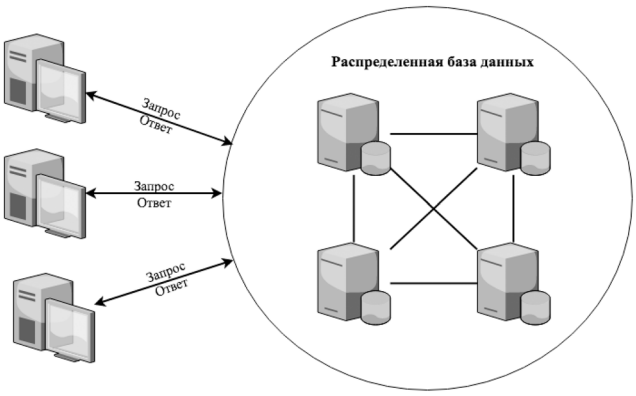

Рис. 1. Архитектура сети типа клиент-сервер для распределенной базы данных

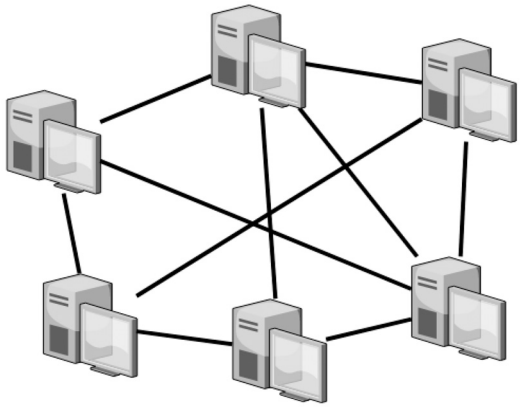

Рис. 2. Одноранговая пиринговая сеть[3]

ционная база данных, в свою очередь, представляет собой набор взаимоувязанных между собой некоторыми отношениями нескольких таких таблиц (сущностей). Такой вид баз данных основан на реляционной модели данных — логической модели данных, сформулированной еще в 1970 г. британским специалистом в области информатики и вычислительной техники Э. Ф. Коддом [2]. Реляционные базы данных отличаются высокой степенью централизации: любые операции с данными (запрос, изменение, добавление, удаление и проч.) обрабатываются единым центром (процессором, CPU), сервера которого, чаще всего, физически расположены в одном месте и администрирование которых осуществляется неким специальным лицом/организацией. Пользователи в такой системе работают с данными в формате «запрос – ответ»[1].

Другой крупный вид баз данных — распределенная база данных (distributed database, DDB), которая, по сути, представляет собой сеть из нескольких взаимосвязанных баз данных (нодов), распределенных в компьютерной сети. Любые операции с данными в такой базе обрабатываются децентрализовано — сетью из нескольких центров (процессоров, CPU), при этом данные распределены по различным хранилищам (серверам) и даже могут частично дублироваться. С развитием интернета, потребности компаний в хранении и обработке большого количества структурированных и неструктурированных данных росла, а распределенные базы данных оказались наиболее подходящими в плане повышенного уровня отказоустойчивости (отсутствие единой точки отказа, выход из строя которой повлечет за собой неработоспособность всей базы) и масштабируемости (удаленная аренда облачного сервера обходиться дешевле и выгоднее для некоторых видов бизнеса, чем покупка нового физического сервера) вариантами[2].

Все перечисленные базы, так или иначе, построены на архитектуре «клиент-сервер» — клиенты (пользователи) посылают некоторые запросы на чтение или редактирование данных, а сервера, объединенные в распределенную сеть, их исполняют, храня эти данные у себя (см. рис. 1).

Стоит также отметить, что существует и другой вид распределенной сети, называемый одноранговой пиринговой сетью (peer2peer network), в которой могут отсутствовать выделенные сервера, так как их функции выполняют сами клиенты сети (см. рис. 2).

Таким образом, можно выделить несколько важных особенностей традиционных и распределенных баз данных. Первой особенностью таких баз является централизация, так как за их функционирование отвечает, обычно, один или несколько центров ответственности. Предполагается, что эти центры действуют в условиях полного доверия друг к другу, а также пользуются доверием со стороны пользователей. Более того, на плечах центра лежит ответственность за проведение и валидацию (одобрение или отказ) транзакций — последовательности операций над базой (например, внесение или корректировки данных), переводящих базу из одного целостного состояния в другое.

Вторая особенность — это отражаемость данных в текущий момент времени. Каждый раз, когда клиент в некоторый момент времени обращается к базе, он видит ее текущее состояние, причем чаще всего без возможности увидеть данные в том виде, в котором они были вчера, неделю назад или год назад. Конечно же, современные базы порой предлагают возможности просмотра истории изменения того или иного элемента (так называемые персистентные структуры данных). Администраторами баз на регулярной основе создаются резервные копии (бэкапы), представляющие собой набор «слепков» базы в некоторый момент времени.

---

[1] Для работы с такими базами обычно используется язык программирования SQL. Самыми популярными реляционными системами управления баз данных являются решения, предоставляемые компаниями Oracle (Oracle Database), IBM (IBM DB2), Microsoft (Microsoft SQL Server) и проч.

[2] Известными распределенными базами данных являются: распределенные SQL базы (от Microsoft, Oracle, IBM); нереляционные NoSQL базы (MarkLogic, MongoDB.); NewSQL (Google Spanner, Clustrix), объединяющие в себе первые два подхода; проект Hadoop, используемый для обработки и хранения «больших данных».

[3] Самая известная реализация такой сети — протокол BitTorrent, позволяющий пользователям предоставлять различные файлы для скачивания другим пользователям. Файлы реплицируются по сети, и пока хоть один из владельцев участвует в процессе обмена, файл остается доступен для других

Тем не менее, в таких базах совершенно отсутствует механизм верификации того, что информация не была изменена задним числом, что приводит к проблеме доверия к администратору.

Третья особенность — возможность клиентом выполнять базовые операции с данными, обозначаемых акронимом CRUD (create, read, update, delete) — создавать записи в базе, читать их (возможность просматривать), обновлять и удалять их.

Место распределенного реестра в иерархии видов баз данных представляется не совсем очевидным. В первую очередь это связанно с путаницей терминов база данных (database) и реестр (ledger), которые чаще всего употребляются как синонимы (см. например [3, 4][4]). Некоторые эксперты определяют распределенный реестр как один из видов распределенной базы данных [5].

Распределенный реестр, на сегодняшний день, определяют чаще всего как некую базу данных, распределенную в компьютерной сети между различными центрами (нодами). В отличие от классической распределенной базы данных, в распределенном реестре предполагается хранение всей полной и актуальной базы на каждом из нодов. Данная, на первый взгляд, избыточность обусловлена другой характеристикой распределенного реестра — отсутствием доверия пользователей друг к другу или к центру (см. [6]). Такие условия могут возникнуть, например, в случае, когда у пользователей базы данных имеются основания полагать, что центр, ее администрирующий, имеет возможность манипулировать данными, нарушая их целостность и искажая информацию в них заложенную. Иными словами, база данных, построенная по технологии распределенного реестра, представляет собой одноранговую пиринговую сеть. Следовательно, вся ответственность за правильность вносимой и отражаемой в такой базе информации равномерно распределяется между всеми участниками. Отсюда возникает потребность в создании некоторого механизма согласования, с помощью которого все участники сети в условиях неполного доверия друг к другу будут иметь исправно функционирующую и адекватно отражающую текущее состояния данных базу. Такой механизм носит название механизм (алгоритм) консенсуса.

Механизм консенсуса представляет собой некий компьютерный алгоритм, который лежит в основе распределенного реестра. На сегодняшний день существует множество механизмом консенсуса. Выбор конкретного механизма обосновывается целями создания реестра, а также природой активов, в нем отраженных. Задачей механизма консенсуса является определение легитимности (корректности) каждой производимой в базе транзакции и вынесение вердикта о возможности ее проведения, с использованием заранее определенного метода криптографической валидации, принятого в данном распределенном реестре. Данный механизм автоматизирует процедуру выработки единого мнения среди участников сети относительно корректности отражения данных на текущий момент времени в распределенном реестре.

Другой важной задачей механизма консенсуса является разрешение конфликтов между некоторыми противоречащими транзакциями, проводимыми одновременно (т. н. проблема двойного расходования). Например, операция над каким-либо активом в базе, инициированная в один и тот же момент разными нодами (узлами). Механизм коненсуса, таким образом, обеспечивает последовательное выполнение всех транзакций, а также выполняет защитную функцию от захвата контроля над распределенным реестром группой лиц (нодов), для проведения некорректных (нелегальных, обеспечивающих, например, двойное расходование) транзакций (особенно, в случае общедоступного распределенного реестра).

Отличительной чертой распределенного реестра является активное использование криптографических методов. В первую очередь, речь идет о криптографически стойкой хэш-функции (hash-function), представляющую собой одностороннюю функцию $h\,(k)$, которая преобразовывает массив входных данных произвольной длины k в битовую строку установленной длины. Например, подав на вход 128-битного алгоритма хэширования MD5 k=«привет», в результате получим

$$h\,(k)=608333adc72f545078ede3aad71bfe74.$$

Если в исходном k будет изменен хотя бы один байт, то $h\,(k)$ примет кардинально иное значение. Например, пусть k=«превет», тогда значение

$$h\,(k)=f75c47b64316ce38ac04ec72b0ae0a98.$$

Хэш-функции обладают двумя важными свойствами. Во-первых, зная выходное значение функции, входное значение k идентифицировать вычислительно сложно за любое мыслимое время. Во-вторых, функция при разных значениях аргумента принимает разные значения (не существует двух значений аргумента, хэш-функции от которых одинаковы).

Другим важным технологическим аспектом, связанным с криптографией в распределенном реестре, является использование пары цифровых закрытого и открытого ключей. Открытый ключ является производным от закрытого, причем создается также с помощью некоторой односторонней криптографически стойкой функции. Данная пара ключей служит аналогом обычной подписи на физических документах и выполняет, по сути, ту же самую функцию. Продолжая аналогию с физической подписью, можно условно сравнить открытый ключ с тем, как подпись выглядит на бумаге, а закрытый ключ — непосредственно с рукой подписанта. Наблюдая открытый ключ, получатель имеет все основания полагать, что входящее сообщение/транзакция было подписано именно с помощью закрытого ключа определенного подписанта.

Еще одной важной особенностью распределенного реестра является хранение всей истории

[4] Использование термина реестр (ledger) обусловлено, скорее всего, схожестью с термином главной бухгалтерской книги (general ledger), как местом записи всех финансовых операций компании.

транзакций. В отличие от традиционных баз данных, в которых основной единицей учета является актуальное значение какого-либо атрибута, единицей учета распределенного реестра является транзакция (операция изменения некоторого атрибута). Имея в распоряжении всю историю транзакций, можно узнать текущее значение того или иного атрибута. В связи с этим распределенные реестры считаются необратимыми базами, так как изменение уже завершившихся транзакций, имеющих подписанное цифровой подписью определенное значение хэш-функции, невозможно. Хэш от подписанной транзакции, которая, например, представляет собой договор о передачи какого-либо актива от одного лица другому, наблюдается всеми участниками сети и одобряется механизмом консенсуса, действующим в данном распределенном реестре. Раз одобренная транзакция не может быть изменена, так как изменения хотя бы одного символа в свойствах транзакции приведет к кардинальному пересчету значения хэш-функции. Так как каждый нод хранит у себя копию базы, то такое изменение станет известно остальным участникам. Аналогичным образом невозможно провести и процедуру удаления или отмены транзакции, если она уже однажды была одобрена другими нодами в процессе работы механизма консенсуса. Следовательно, любое изменение базы данных распределенного реестра может быть осуществлено лишь с помощью новой транзакции, удаление же или внесение правок, особенно в открытых публичных реестрах, невозможно.

Стоит также отметить, что информация в распределенном реестре, в большинстве случае, носит псевдоанонимный характер, особенно в случае открытого децентрализованного реестра. Так как все участники могут просматривать всю базу транзакций, то зная, какое лицо стоит за тем или иным неким идентификационным номером (в сети биткоин, например, это номер кошелька), любой может проследить историю транзакций интересуемого лица.

2. Блокчейн

Разобравшись с основными отличиями распределенного реестра от традиционных баз данных и определив его отличительные черты, рассмотрим теперь самый первый и до сих пор самый популярный вид распределенного реестра — блокчейн, а конкретно блокчейн биткоина как самый показательный пример распределенного реестра.

Блокчейн, как следует из самого названия, представляет собой цепочку блоков. Каждый блок содержит в себе набор из совершенных в течение определенного периода времени (в биткоине это в среднем 10 минут) транзакций. Транзакции в сети биткоин представляют собой записи, которые фиксируют передачу какого-либо количества биткоинов от одного пользователя другому. Соответственно, транзакция считается одобренной и выполненной, если она включена в какой-то блок. Всего размер блока составляет 1 Мб, средний размер записи о транзакции — 495 байт, то есть 1 блок содержит запись примерно о двух тысячах транзакций.

Стоит отметить, что в блок заносится не сама транзакция, а некоторое 32-битное значение, обозначаемое в биткоине как корень Меркла (merkle_root). Данное значение высчитывается с помощью метода построения дерева Меркла (Merkle tree), представляющего собой процесс построения древовидный иерархии, в котором каждый новый уровень является значением хэш-функции от каких-то двух хэшей предыдущего уровня. На самом нижнем уровне дерева находятся значения хэш-функции от информации, содержащихся в самой транзакции (хэш от записи вида «А передал Б 10 биткоинов»). На один уровень выше высчитывается значение хэш-функции от хэшей двух транзакций. Таким образом, на самом верхнем уровне оказывается некоторый хэш, который и является тем значением корня Меркла, заносимым в блок.

Помимо хэша от набора транзакций, блок содержит так называемый заголовок (block header), уникальный для каждого блока. Его уникальность достигается за счет того факта, что каждый заголовок является зна-

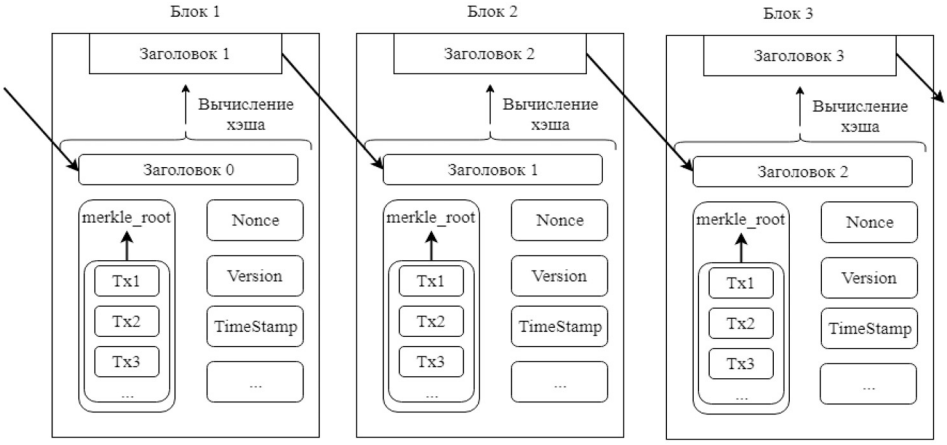

Рис. 3. Схема блокчейна

Примечание. Блоками Tx1, Tx2, Tx3 условно обозначены транзакции, учет которых с помощью построения дерева Меркла дает значение хэш-функции merkle_root (корень Меркла); Nonce — некоторое числовое значение для перебора, используемого для решения вычислительной задачи; Version — текущая версия алгоритма; TimeStamp — временная метка

чением хэш-функции, носящей название SHA256, от информации, хранящейся в нем: списка транзакций (корня Меркла), временной метки создания блока, версии текущего алгоритма, текущей сложности вычисления блока, nonce и др., а также, что важно, от заголовка предыдущего блока. Таким образом, через последовательный учет заголовков предыдущих блоков в вычислении заголовка нового осуществляется их связь, создается цепочка блоков. Схематичное изображение блокчейна представлено на рис. 3.

Блокчейн биткоина представляет собой полностью децентрализованный одноранговый распределенный реестр. Сам блокчейн представляет собой некоторый файл, копии которого находятся на всех нодах одновременно, то есть у всех участников сети, и синхронизируются друг с другом в режиме реального времени. На сегодняшний день блокчейн биткоина занимает 181,05 Гб на жестком диске, а в сети регистрируется чуть более 10 тыс. полноценных нодов, хранящих весь блокчейн биткоина[5]. Так как же происходит процесс добавления новых блоков в цепочку и синхронизация между всеми этим нодами? Ответы на эти вопросы непосредственно связаны с решением задачи о нахождении консенсуса (единого мнения относительно валидности текущего состояния данных, а также легитимности проведения транзакций) в распределенной децентрализованной сети, известной также как задача о византийских генералах.

2.1. Задача о византийских генералах

Данная задача восходит к работе [7], посвященной анализу надежности децентрализованной системы. Авторы формализовали проблему в следующем виде.

Допустим, византийская армия осаждает некоторый вражеский город. Армия разделена на некоторые боевые единицы (полки, дивизии), каждой из которых командует свой генерал. К утру генералы должны прийти к общему решению относительно дальнейших действий (штурм или отступление). Однако среди генералов могут оказаться предатели, которые будут мешать выработке единого решения относительно плана на утро. Следовательно, существует 3 возможных исхода: если все верные генералы атакуют, то город будет взят; если все верные генералы отступят, то армия останется целой; если некоторые верные атакуют, а другие верные отступят — армия будет разгромлена. Итоговое решение достигается посредством обмена информации между генералами относительно состояния своего войска, а также оценки вражеских сил посредством некоторого условного «мессенджера» (например, голубиной почты или гонцов). Таким образом, возникает задача определения, сколько максимально в армии может быть генералов предателей, чтобы армия все равно пришла к единому мнению?

В той же работе [7] было показано частное решение такой задачи с помощью определенного рекурсивного механизма, положенного в дальнейшем в основу семейства протоколов Paxos для достижения консенсуса в сети ненадежных вычислений. В итоге, авторы получили результат, что единый консенсус будет найден в случае, если предателей среди генералов строго меньше одной трети. Стоит отметить, что в том виде полученный результат был, скорее, теоретическим и трудно реализуемым на практике.

Следующий виток развития в дискуссии о решении такого рода задачи был дан работой [8], в которой был теоретически описан первый практически реализуемый алгоритм, устойчивый к проблеме византийских генералов (Practical Byzantine Fault Tolerance Algorithm, PFBT). Данный алгоритм способен обеспечить передачу огромного массива прямых сообщений между участниками в одноранговой сети с минимальной задержкой. Это открыло возможности практической реализации алгоритма в различных компьютерных приложениях, в том числе и в качестве механизма консенсуса в некоторых современных распределенных реестрах.

Тем не менее, первым реализованным на практике механизмом консенсуса, который в условиях открытой одноранговой сети в форме распределенного децентрализованного реестра смог обеспечить устойчивость к проблеме византийских генералах, стал алгоритм, используемый в блокчейне биткоина.

2.2. Доказательство работы

Механизм консенсуса, используемый в блокчейне биткона, носит название «доказательство работы» (Proof-of-Work). Суть его заключается в том, что полноценные ноды участвуют в своего рода соревновании за первенство генерации каждого нового блока. Оно также носит названия майнинга (mining, добыча). Наградой в таком соревновании является некоторое число биткоинов (на сегодняшний день — 12,5 + комиссии от транзакций). Само по себе состязание заключается в попытке подобрать некоторый хэш определенной сложности, который будет служить заголовком нового блока. Сложность определяется самим алгоритмом в зависимости от совокупной вычислительной мощности всех нодов и корректируется каждые 2016 блоков (раз в две недели). Это делается для того, чтобы поддерживать на стабильном уровне частоту добавления нового блока в цепь (раз в 10 минут). Соответственно, будь в сети биткоин 5 участников с обычными домашними персональными компьютерами или миллион суперкомпьютеров, — блоки все равно будут генерироваться в среднем раз в 10 минут.

Сама по себе сложность представляет собой определенное правило того, как должен выглядеть хэш, служащий заголовком блока. Сам хэш представляет собой набор из 64 буквенных и числовых символов, а сложность регулирует количество нулей в его начале. Например, в 2010 г. начало заголовка должно было содержать 8 нулей, а в 2018 г. — уже 19, что говорит о значительном приросте мощностей в сети биткоин. Типичный заголовок блока выглядит так: «00000000 00000000000fc05c87cb16c87e2988984a60332d209b92 8a7 4d8bff5».

[5] Большая часть из 11 млн пользователей данной криптовалюты используют специальные сайты и легковесные клиенты, позволяющие пользоваться биткоином без скачивания непосредственно базы.

Сам процесс подбора хэша заключается в следующем. У участника сети, желающего помайнить, по большому счету, есть две переменные в свойствах блока, которые он может варьировать с целью подбора заголовка нужного вида: это корень Меркла и Nonce. В зависимости от того, какие транзакции будут включены в блок, корень Меркла будет принимать разные значения, что, как видно на выше приведенном рис. 3, влияет на хэш заголовка. Тем не менее, стандартной практикой является набор в блок транзакции из списка ожидания с самой большой комиссией, которая также является прибавкой к вознаграждению майнеру за генерацию блока. Основной процесс майнинга, следовательно, заключается в простом переборе числового значения Nonce до тех пор, пока заголовок блока не будет выглядеть надлежащим образом.

После выполнения всех необходимых условий, майнер публикует блок с указанием всех необходимых атрибутов, включая найденное значение Nonce. Зная все атрибуты, полные ноды в автоматическом режиме перепроверяют, действительно ли при таких входных данных и таком значении Nonce, хэш заголовка будет выглядеть именно так, а не иначе. После подтверждения майнеры переключаются на генерацию нового блока, а автор только что созданного получает на свой электронный биткоин-кошелек вознаграждение, причем награда будет доступна для траты лишь после того, как от его блока будут сгенерированы и подтверждены еще более 100 блоков.

Вполне возможно представить ситуацию, когда два майнера одновременно сгенерировали подходящие блоки, что вызывает раздвоение цепи. Тогда главной цепочкой будет считаться та, ответвление которой будет быстрее продолжено (см. рис. 4).

На рис. 4 белым отмечен исходный блок, черным — главная цепочка, темно-серым — ответвления. Считается, что транзакция считается окончательно проведенной, если от блока, в котором она учтена, цепь блоков продлена хотя бы еще на 6 блоков. В противном случае, если транзакция по передачи некоторого количества биткоинов оказалась в отвергнутой цепочке, она становится недействительной и автоматически вновь становится в очередь на подтверждение при генерации текущего блока. Соответственно, майнер, сгенерировавший отвергнутый блок в блокчейне Биткоина, никакого вознаграждения не получает.

Доказательство работы считается относительно эффективным в плане предотвращения двойного расходования механизмом консенсуса. В централизованных системах создание двух транзакций, например, передающих один и тот же актив двум разным агентам невозможно в силу наличия центра, не позволяющего проведения такого рода транзакций. В блокчейне биткоина при подборе транзакций в блок в случае, когда один и тот же биткоин был отправлен двум разным адресатам, возникнет конфликт, и в блок попадает лишь одна из этих транзакций. Тем не менее, существует теоретическая возможность того, что двойное расходование может быть проведено, если какое-то лицо или группа лиц будет контролировать более 50% вычислительных мощностей сети (т. н. «атака 51%»), хотя на практике в сети биткоин этого никогда не происходило[6].

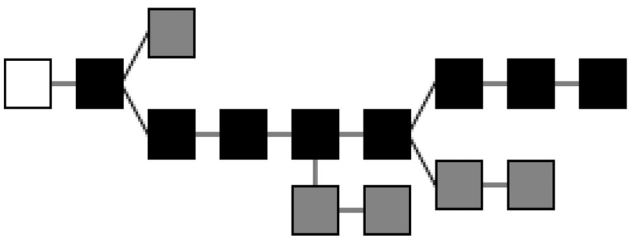

Рис. 4. Случай раздвоения цепочки блоков

С другой стороны, потенциальной уязвимостью, с которой также хорошо справляется доказательство работы, является возможность изменения задним числом уже одобренной транзакции. При изменении хотя бы одной транзакции в блоке, значение хэш-функции в заголовке данного блока поменяется, что повлечет изменение всех последующих заголовков блоков данной цепи. Так как изменение в транзакцию вносится каким-то отдельным лицом, то его цепь будет игнорироваться сетью, так как полученные заголовки будут не соответствовать правилу о нескольких нулях в начале. Единственное, как возможно распространить скорректированную версию блокчейна — это пересчитать по новой все хэши, чтобы они удовлетворяли заданным условиям. Это, однако, невозможно, так как займет очень много времени, и злоумышленник просто не сможет догнать сеть, которая как минимум с такой же скоростью, с которой прошлый блок переписывается, будет добавлять в цепь новый.

На сегодняшний день консенсус «доказательство работы» зарекомендовал себя как самый надежный и защищенный механизмом консенсуса для открытого, публичного и децентрализованного распределенного реестра в форме блокчейна (см., например, [9]). Тем не менее, несмотря на все плюсы, данный алгоритм имеет и ряд недостатков.

В первую очередь, недостатком доказательства работы называют его высокую энергозатратность. Например, по данным исследования [10] прогнозируется, что к концу 2018 г. затраты на майнинг только лишь биткоинов достигнут 0,5% (7,67 ГВт) от общего энергопотребления планеты. В настоящий момент на майнинг данной криптовалюты приходится порядка 2,55 ГВт электричества, что сопоставимо с потреблением электричества Ирландией (3,1 ГВт), а в Исландии на майнинг уже тратится больше электричества, чем потребляют обычные граждане страны.

Вторая главная проблема доказательства работы, это скорость фиксируемых в системе транзакций. Обычно, данный аргумент используется в рамках дискуссии о перспективах криптовалют в качестве полноценных платежных средств. Самые популярные криптовалюты биткоин и эфир (Ethereum) могут проводить 7 и 20 транзакций в секунду соответственно,

[6] Единожды в 2014 г. пул майнеров (объединение майнеров, которые решают задачу нахождения хэша посредством совокупной распределенной вычислительной мощности каждого участника пула) Ghash.io получил контроль над 55% мощностей сети, что вызвало падение курса криптовалюты на 25%, а пул обязался снизить свои мощности до 40%. Тем не менее, как таковой атаки не произошло.

плюс время ожидания на подтверждение нескольких следующих блоков в цепи для полной уверенности в проведении платежа. Для сравнения, централизованная система Visa проводит 24 тыс. транзакций в секунду. С точки зрения платежных средств, такая медлительность децентрализованных реестров в виде блокчейна играет значительную роль. Если же абстрагироваться от криптовалют и говорить о построении некоторой базы данных на блокчейне, то такая скорость фиксирования транзакций может быть вполне приемлемой.

Блокчейн биткоина, использующий доказательство работы в качестве механизма консенсуса, является полностью децентрализованным и открытым. Открытость заключается в том, что никто не контролирует доступ для участия в процессе обмена биткоинами, а также никто не ограничивает доступ к чтению всей базы транзакций. Любой пользователь может свободно скачать весь блокчейн, создать учетную запись (кошелек) и начать майнить биткоины. Для такого открытого и децентрализованного реестра механизм доказательства работы является лучшим вариантом, так как гарантирует надежность данных.

Надежность и открытость привлекает множество организаций к работе с блокчейном биткоина не только как с сетью для передачи криптовалюты, но и как с методом сохранения некоторой информации, которая может быть записана в метаданные (свойства) транзакций. Именно эту особенность использует платформа Exonum, предоставляющая возможность создания приватных блокчейнов с функцией записи их состояний в блокчейн биткоина. Иными словами, в метаданные транзакции по передаче минимального количества биткоинов записывается хэш от состояния некоторого приватного блокчейна в определенный момент времени. Именно так работают земельные кадастры Украины и Грузии, что исключает возможность подлога и изменения документов задним числом.

Тем не менее, большое количество имплементаций технологии распределенного реестра в различных областях экономики не отвечают принципам открытости и децентрализации, которые воспринимаются многими криптоэнтузиастами в качестве необходимых условий функционирования таких систем. Однако внедрение блокчейна даже с нарушением данных принципов может значительно повлиять на эффективность ведения бизнеса. От целей, для которых создается распределенный реестр, зависит степень его открытости и централизации. Данные параметры, в свою очередь, влияют на выбор оптимального механизма консенсуса. В связи с этим представляется важным представить классификацию видов распределенных реестров, а также альтернативные механизма консенсуса.

3. Виды распределенных реестров

Многие эксперты и исследователи предпринимали попытку классификации видов распределенного реестра (см. например [11-14]). Несмотря на употребление чуть разных терминов, в целом все исследователи сходятся к некоторому общему мнению относительно их классификации.

По степени открытости распределенные реестры могут быть открытые (permissionless) и закрытые (permissioned). Открытый реестр не имеет механизма отбора участников, т. е. любой пользователь может присоединиться к сети и не может быть никоим образом ограничен в доступе к ней. В закрытом реестре доступ предоставляется лишь тем лицам, кто удовлетворяет некоторым заранее заданным требованиям и/или чья кандидатура одобряется владельцем или администратором реестра. Закрытый реестр, таким образом, является довольно привлекательным вариантом для правительства и регуляторов, так как дает возможность однозначно идентифицировать участников и контролировать доступ к сети, например, посредством процедуры регистрации или выдачи лицензий. Такие реестры также легче поддаются регулированию. Обратной стороной, конечно же, является жертва главным преимуществом распределенного реестра — способностью функционировать без какого-либо координационного центра. Тем не менее, закрытый распределенный реестр не обязательно предполагает наличие центрального органа, контролирующего проведение всех транзакций — ограничение может быть только на вход, тогда как внутри реестра все участники все равно остаются в равном положении.

По степени централизации распределенные реестры бывают децентрализованными (публичными, public), частично централизованными (федеративными, federated) и полностью централизованными (частными, private). Абсолютно децентрализованный распределенный реестр предполагает, что валидацией транзакций могут занимаются все участники сети. Также такие реестры обычно являются открытыми. Если, однако, такой реестр имеет ограничения на вход, то есть является закрытым, и/или валидацией транзакций занимается лишь ограниченное число нодов (т. е. список таких нодов представляется относительно закрытым), то распределенный реестр считается частично централизованным. Закрытый частично централизованный распределенный реестр также носит название «консорциум» (consortium). Если же валидацией транзакций в реестре занимается отдельный выделенный центр, то такой реестр является полностью централизованным.

Таким образом, приведенные выше идеи можно формализовать с помощью следующей схемы (см. рис. 5).

Также следует отметить, что современные блокчейн-платформы поддерживают смарт-контракты (smart-contracts) — компьютерные коды, позволяющие осуществлять определенные транзакции в распределенном реестре при наступлении тех или иных заранее запрограммированных событий. Данный аспект открывает широкое окно возможностей в мире различных финансовых и банковских продуктов, так как смарт-контракт способен облегчить или даже заменить посредника в виде банка, например, при использовании аккредитивов, векселей или эскроу счетов, а также обеспечить большую прозрачность в проведении такого рода сделок. Логично предположить, что распределенные реестры для таких целей должны быть закрытыми для сохранения коммерческой тайны.

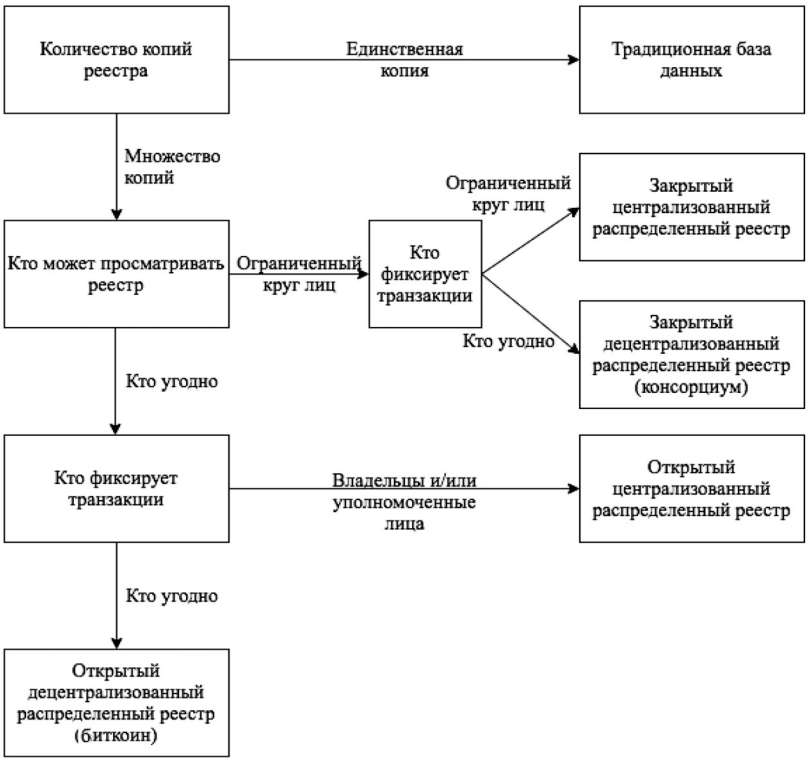

Рис. 5. Виды распределенного реестра по степени открытости и централизации

Описанный ранее алгоритм консенсуса в форме доказательства работы, используемый в биткоине, очевидно является избыточным для некоторых видов распределенного реестра. В закрытом централизованном реестре нет смысла некоторому уполномоченному лицо тратить большие вычислительные мощности только для формирования блоков. Это лицо ни с кем не конкурирует за награду и имеет возможность, при желании, даже переписать информацию во всей цепочки блоков, так как оно единственное занимается майнингом. Конечно, эти изменения могут быть зафиксированы другими участниками в сети. Тем не менее, сам механизм майнинга в данном случае является излишним.

В связи с этим, а также с попыткой решить описанные выше проблемы механизма доказательства работы, свет увидели другие алгоритмы консенсуса. Какие-то из них лучше всего применимы лишь к отдельным видам распределенного реестра с учетом открытости и централизации, другие являются более универсальными. Разнообразие консенсусов и типов распределенного реестра уже сегодня породили совершенно новый рынок «блокчейна-как-услугу» (Blockchain-as-a-Service, BaaS), на котором различные стартапы, а также такие крупные гиганты как IBM и Microsoft, предлагают различные платформы для создания блокчейнов в зависимости от требований корпоративных или государственных заказчиков.

3.1. Доказательство владения и его производные

Вторым по популярности алгоритмом консенсуса на сегодняшний день является механизм доказательство владения (proof-of-stake, другое название — подтверждение доли). Основным мотивом его появления послужила обеспокоенность широкой общественности огромными вычислительными мощностями, которые тратятся «впустую» при майнинге криптовалют, основанных на доказательстве работы. В итоге, в данном алгоритме консенсуса доминантой оказывается количество криптовалюты/токенов на балансе отдельного участника. Если при использовании доказательства работы вероятность подобрать заголовок следующего блока зависит от располагаемой вычислительной мощности, то при использовании доказательства владения вероятность зависит от баланса пользователя. Формирование блоков с помощью механизма доказательства владения осуществляется значительно быстрее, что увеличивает общую пропускную способность распределенного реестра. Сам процесс генерации блоков носит название форджинг (forging — ковка). При этом в процессе создания блока не происходит добыча нового количества криптовалюты — в качестве вознаграждения выступает лишь комиссия от включенных в блок транзакций[7].

Очевидно, что возможна ситуация, в которой большую часть монет в системе будет контролировать одно лицо, что приведет к ее централизации. Такое лицо будет иметь возможности манипулировать информацией в блоках и получит контроль над сетью. Тем не менее, предполагается, что злоумышленнику это будет невыгодно, так как приведет к неустойчивости сети, падению доверия к валюте, понижению ее курса, что, в первую очередь, ударит по тому, кто

[7] Количество монет, таким образом, может быть фиксировано и распределено заранее, по аналогии с первичным размещением акций. Возможен также вариант комбинированного использования механизма доказательства владения и работы, когда некоторая часть блоков майнится, чем обеспечивается дополнительные вливания криптовалюты в обращение.

владеет самым большим ее количеством. Впрочем, сама ситуация, когда больше 51% валюты находится в руках одного лица является труднодостижимой: для этого злоумышленнику придется закупать монеты у других пользователей, что приведет к росту курса и повышению затрат на ее приобретение.

Также в некоторых криптовалютах[8] используется чуть видоизмененный вид доказательства владения — делегированное доказательство владения (Delegated Proof-Of-Stake, DPoS). Суть заключается в том, что держатели могут использовать свои средства как инструмент голосования для выбора тех, кто будет заниматься генерацией блоков. Например, на блокчейн-платформе EOS существует 20 избираемых должностей валидаторов транзакций, еще 1 отбирается случайным образом. Пул валидаторов обновляется каждую минуту, а блок принимается, если его одобрили 50+1% валидаторов. Выбор того, кто конкретно генерирует блок определяется некоторым псевдорандомным алгоритмом. Причем на вероятность быть отобранным одним из 21 валидаторов для генерации следующего блока не зависит от количества монет на счету и времени, проведенным в качестве валидатора и проч. Стоит отметить, что, несмотря на кажущуюся сложность, алгоритм работает автоматически, почти моментально, а новые блоки генерируются каждые 1-3 секунды, обеспечивая высокую скорость проведения транзакций. При этом обычный участник может в любой момент отдать свой голос другому кандидату в валидаторы. Таким образом, за счет наличия системы голосования предотвращается возможность захвата блокчейна узким кругом лиц. С другой стороны, будучи в своей природе демократическим механизмом, не исключена возможность сговора между владельцами большого количества токенов и делегатами, своего рода лоббизм интересов крупного капитала через избранных представителей.

Интересным также представляется механизм консенсуса доказательства авторитета (Proof-of-Authority), при котором валидаторами становятся заранее установленный круг лиц, своего рода администраторы сети. Например, на блокчейн-платформе POA Network, первой реализовавшей такой алгоритм, стать валидатором можно лишь пройдя предварительную процедуру лицензирования с предоставлением большого количества персональных данных, которая нотариально заверяется в США. Валидатор в таком варианте становится публичным лицом. По большому счету, такой механизм хорошо подходит для закрытых распределенных реестров в условиях почти полного доверия между участниками.

3.2. Механизмы консенсуса на основе решения задачи о византийских генералах

Данные механизмы представляют собой непосредственную имплементацию существующих частных решений задачи о византийских генералах в распределенные реестры. Одной из особенностей таких механизмов консенсуса является ограниченная масштабируемость, однако при этом достигается высокая скорость проведения транзакций. Если при использовании доказательства работы большое количество нодов (валидаторов) является, скорее, плюсом, так как гарантирует децентрализацию, и с увеличением их числа скорость обработки транзакций не падает, то вычислительная сложность решения задачи о византийских генералах значительно возрастает при увеличении количества генералов, что отрицательно сказывается на времени выработки консенсуса и скорости проведения транзакций. В связи с этим, данный тип механизма консенсуса чаще применяется в закрытых корпоративных распределенных реестрах.

По сути, решение задачи о византийских генералах сводится к обмену информацией между генералами, сравнения полученных данных и выработке общего мнения. Именно эта процедура и используется в

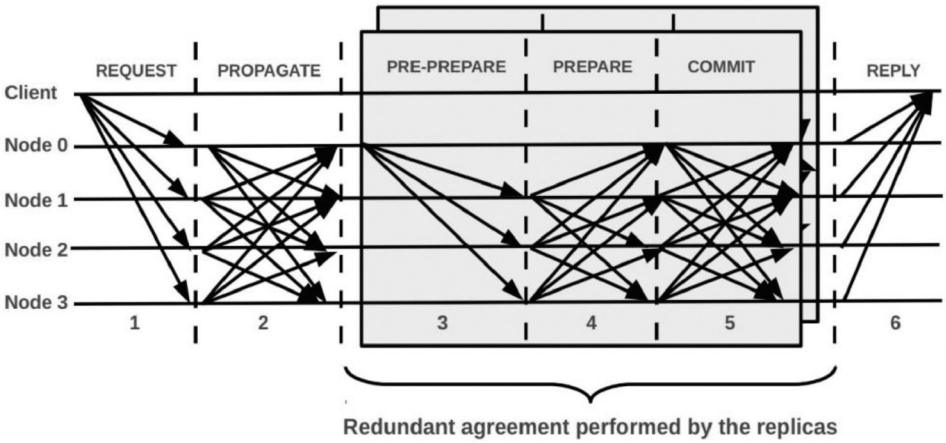

Рис. 6. Схема работы Hyperledger Indy

Примечание. Одобрение транзакции происходит в 6 этапов. Первый этап — запрос, который клиент отправляет нодам; второй — распространение; третий — преподготовка; четвертый — подготовка; пятый — соглашение; шестой — принятие. Избыточное соглашение осуществляется репликами на этапах с третьего по пятый.

Источник: [15]

[8] EOS, BitShares, Steem и проч.

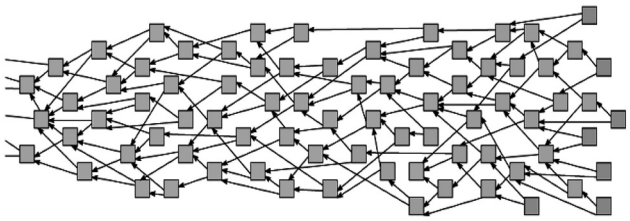

Рис. 7. Направленный ациклический граф
Источник: [16]

рамках различных блокчейн-платформ проекта Hyperledger[9].

На рис. 6 приведена иллюстрация работы, так называемого, избыточного алгоритма консенсуса, устойчивого к проблеме византийских генералов (Redundant Byzantine Fault Tolerance), используемого в рамках платформы Hyperledger Indy, созданной для построения децентрализованных систем идентификации личности в учреждениях, организациях, а также между ними.

На первом этапе клиент посылает запрос о проведение какой-либо транзакции некоторому числу заранее избранных за рамками системы нодам (валидаторам). Для эффективности, запрос достаточно распространить f+1 нодам, где f — предполагаемое количество потенциально нечестных валидаторов в сети. На втором этапе информация о поступившем запросе распространяется между всеми нодами в сети. Третий этап, преподготовка, заключается в том, один из нодов, называемый «мастером», инициирует процедуру принятия, т. е. утверждает, что транзакция валидна и выдвигает предложение другим нодам о проверки ее валидности со своими копиями баз (репликами). На четвертом этапе ноды рассылают информацию о текущем состоянии своих реплик и сравнивают реплики других со своими с учетом новой транзакции. На пятом этапе ноды обмениваются информацией о согласии с внесением транзакции в реестр, что и происходит на шестом этапе, вместе с сообщением клиенту результатов обработки его запроса.

Помимо корпоративных решений, механизмы консенсуса, основанные на решение задачи о византийских генералах, используются также и в некоторых публичных блокчейнах и криптовалютах. В Ripple, позиционирующей себя как международное платежное средство, реализован так называемый Ripple Protocol Consensus Algorithm (RPCA), который подразумевает наличие некоторого множества заранее отобранных компанией Ripple Foundation валидаторов (сейчас их 50), среди которых сами разработчики, криптовалютные биржи, крупные технологические стартапы, а также несколько крупных компаний (шведский оператор сотовой связи Bahnhof, Microsoft и др.) и Массачусетский технологический институт.

Стоит отметить, что несмотря на кажущуюся многоаспектность и сложность, алгоритмы консенсуса, основанные на решение задачи о византийских генералах, выполняются автоматически без каких-либо усилий со стороны пользователя. Скорость проведения и фиксирования транзакций при этом превышают показатели биткоина и эфира в тысячи раз[10].

3.3. Направленный ациклический граф

Говоря о видах распределенных реестров, до этого момента мы в основном имели ввиду блокчейн, как самую распространенная форма реестра. Тем не менее, относительно недавно появился новый вид распределенного реестра, получивший название направленного (ориентированного) ациклического графа (Directed Acyclic Graph, DAG).

Подход, реализованный в распределенных реестрах на базе направленного ациклического графа концептуально отличается от реестров на блокчейне. В отличии от обычного блокчейна, где транзакции «складываются» в блок, в данном случае этого не происходит и остается лишь связь транзакций друг с другом напрямую. Иными словами, можно вообразить некоторый ориентированный граф, где вершины — это транзакции, а ребра — связи между транзакциями через хэш-функцию. Так, собственно, и выглядит данный вид распределенного реестра (см. рис. 7).

Данную архитектуру распределенного реестра можно описать также как взаимосвязанную совокупность блокчейнов, где каждый блок содержит лишь одну транзакцию. Хотя, конечно, такое сравнение не совсем корректно, так как стандартные блокчейны всячески стараются предотвратить выстраивание параллельных цепей, что является совершенно нормальным для ориентированного ациклического графа.

Одной из реализаций данной концепции является механизм Tangle криптовалюты IOTA, созданной для торговли данными в интернете вещей. Этот проект позволяет торговать данными, собираемые любыми «умными» устройствами. Для того чтобы провести транзакцию в сети, использующей Tangle, пользователь (не сам пользователь, конечно же, но его клиентское приложение) должен подтвердить несколько других транзакций (в данный момент 2). Предполагается, что ноды проверяют транзакции на конфликтность. Если выясняется, что транзакция не соотносится с историей реестра, например, предпринимается попытка передачи монет/токенов, которых нет в распоряжении пользователя, то такая транзакция отклоняется. Как только транзакция набирает достаточное количество одобрений, то она считается проведенной, для нее высчитывается некоторый Nonce (по аналогии со связями блоков в цепи с использованием алгоритма доказательства работы) и через хэш осуществляется связка с предыдущей транзакцией по передаче этого же актива.

[9] В международный консорциум Hyperledger, созданный по инициативе Linux Foundation, входят такие крупные компании как Intel, IBM, American Express, SAP и др.

[10] Например, при заявленной скорости транзакций в сети Ripple в размере 1500 транзакций в секунду, разработчики заявляют, что в силу относительно простой масштабируемости, систему можно «разогнать» до скоростей, сопоставимых с Visa, то есть до 24 тыс. транзакций в секунду. Уже сегодня платежная система PayPal с 193 транзакциями в секунду значительно проигрывает Ripple.

Теоретически, если в какой-то момент некое лицо сможет генерировать хотя бы треть всех транзакций, то возможна ситуация, когда невалидные транзакции будут одобряться. До тех пор, пока участников сети недостаточно много, создатели держат в сети так называемого «координатора», — определенный узел сети, который еще раз перепроверяет все поступающие одобренные транзакции на валидность, прежде чем окончательно их принимать. Разработчики утверждают, что «координатор» исчезнет, как только сеть станет достаточно большой.

Заключение

Технология распределенного реестра является динамично развивающейся областью, в которой новые решения предлагаются каждый день. Выше были продемонстрированы лишь некоторые, самые основные механизмы достижения консенсуса. В действительности, практически каждый блокчейн-стартап или криптовалюта привносит что-то новое в свои продукты, пытаясь превзойти существующие ограничения в масштабируемости, скорости проведения и фиксирования транзакций и анонимности применительно ко всевозможным областям экономики. Таким образом, на сегодняшний день существует большое количество различных реализаций технологии распределенных реестров разной степени открытости, использующих разные механизмы консенсуса

Блокчейн и технология распределенного реестра в целом на сегодняшний день представляется крайне перспективной с точки зрения хранения данных, а также выстраивания децентрализованного механизма взаимодействия внутри фирмы, некоторого бизнес-сообщества или даже между государствами. Уже сегодня многие крупные компании из разных сфер экономики, консорциумы банков, а также центральные банки и государственные структуры запускают свои пилотные проекты, основанные на технологии распределенного реестра. Тем не менее, стоит понимать, что с точки зрения корпоративного и государственного сектора это всего лишь новый тип устройства базы данных, который, однако, в перспективе способен изменить облик бизнес-процессов, лежащих в основе операционной деятельности фирм, и повысить прозрачность органов власти.

Список использованных источников

1. S. Nakamoto (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf.
2. E. F. Codd (1970). A relational model of data for large shared data banks//Communications of the ACM, 13(6), 377-387.
3. G. Greenspan (2018). R3 Corda: Deep dive and technical review | MultiChain. https://www.multichain.com/blog/2018/05/r3-corda-deep-dive-and-technical-review.
4 V. Cealicu et al. (2016). How is the blockchain different from distributed databases in terms of record keeping? https://www.quora.com/How-is-the-blockchain-different-from-distributed-databases-in-terms-of-record-keeping.
5. S. Meunier (2016). Blockchain technology — a very special kind of Distributed Database. https://medium.com/@sbmeunier/blockchain-technology-a-very-special-kind-of-distributed-database-e63d00781118.
6. R. Brown (2016). On Distributed Databases and Distributed Ledgers. https://gendal.me/2016/11/08/on-distributed-databases-and-distributed-ledgers.
7. L. Lamport, R. Shostak, M. Pease (1982). The Byzantine Generals Problem//ACM Transactions On Programming Languages And Systems, 4 (3), 382-401.
8. M. Castro, B. Liskov (1999). Practical Byzantine fault tolerance// OSDI. Vol. 99. P. 173-186.
9. BitFury Group (2015). Proof of Stake versus Proof of Work. https://bitfury.com/content/downloads/pos-vs-pow-1.0.2.pdf.
10. A. de Vries (2018). Bitcoin's Growing Energy Problem//Joule, 2(5), 801-805.
11. World Bank (2017). Distributed Ledger Technology (DLT) and Blockchain. FinTech Note | No. 1. http://documents.worldbank.org/curated/en/177911513714062215/pdf/122140-WP-PUBLIC-Distributed-Ledger-Technology-and-Blockchain-Fintech-Notes.pdf.
12. V. Shermin, V. Kalinov (2017). Blockchain A Beginner Guide (p. 13-17). https://blockchainhub.net/blockchain-technology.
13. P. Kravchenko (2016). Ok, I need a blockchain, but which one? https://medium.com/@pavelkravchenko/ok-i-need-a-blockchain-but-which-one-ca75c1e2100.
14. B. Jew, G. Samman (2016). Blockchain and Shared Ledgers: The New Age of the Consortium. https://www.gtlaw.com.au/insights/blockchain-and-shared-ledgers-new-age-consortium.
15. Hyperledger Architecture, Volume 1. (2017). https://www.hyperledger.org/wp-content/uploads/2017/08/Hyperledger_Arch_WG_Paper_1_Consensus.pdf.
16. R. Amoros (2018). Transactions Speeds: How Do Cryptocurrencies Stack Up To Visa or PayPal? https://howmuch.net/articles/crypto-transaction-speeds-compared.
17. Z. Witherspoon (2018). A Hitchhiker's Guide to Consensus Algorithms — Hacker Noon. https://hackernoon.com/a-hitchhikers-guide-to-consensus-algorithms-d81aae3eb0e3.

Blockchain and distributed ledgers as a type of database

K. D. Shilov, research fellow.

A. V. Zubarev, PhD in economics, senior research fellow, department for mathematical modelling of economic processes.

(Russian presidential academy of national economy and public administration (RANEPA), Institute of applied economic research)

In recent years, there has been much discussion on distributed ledger technology (DLT) and Blockchain in particular. Many developed countries have launched new projects implementing this technology in different areas of economy. Many organizations aim to understand possible applications of this technology in order to find ways of developing business and optimising current processes to reduce costs. There is an opinion, that DLT has a value comparable to such innovations as the Internet and the Telegraph. The potential of the technology will be fully unfolded in the next 5-10 years, however some pilot Blockchain projects implemented today allow us to talk about new business opportunities and radical optimization of the existing business processes. This supports the hypothesis of a global change in the role of institutions and their place in the economy on the horizon of the next decades due to the massive introduction of DLT. This paper reveals the main principles and some technical features underlying DLT, which are necessary for understanding the possibility of its application in various areas of economy.

Keywords: blockchain, distributed ledger technology, database, cryptocurrency.