DOI: 10.55648/1998-6920-2023-17-3-106-122 УДК 123.456.789

Разработка защищенной системы дистанционного голосования и исследование характеристик её анонимности *

И. С. Дьячкова

Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)

Аннотация: В настоящее время актуальной является задача создания надежной и анонимной системы дистанционного голосования. Существующие системы голосования или не вполне отвечают требованиям анонимности, или являются закрытыми от исследования. Была разработана схема для создания системы анонимного дистанционного голосования, базирующаяся на протоколе «слепой» электронной подписи и алгоритме создания анонимного канала. В данной работе описаны возможные угрозы и уязвимости протоколов, на которых базируется схема построения системы анонимного голосования, в том числе угроза саботажа злоумышленником отправки своего или чужих бюллетеней по каналу связи. Представленная оптимизированная версия алгоритма создания анонимного дистанционного голосования гарантирует невозможность подделки бюллетеней или их повторного использования, а также исключает возможные попытки саботажа пересылки бюллетеней голосующими. Также в статье определены формулы для вычисления оптимального значения величины p – вероятности отправки бюллетеня сразу на сервер. Показано, что значение p необходимо вычислять в зависимости от количества голосующих, чтобы достичь максимальной анонимности голосов участников.

Ключевые слова: анонимное голосование, криптография, анонимная передача данных, информационная энтропия.

Разработка защищенной системы дистанционного голосования и исследование характеристик ее анонимности // Вестник СибГУТИ. 2023. Т. 17, № 3. С. 106–122. https://doi.org/10.55648/1998-6920-2023-17-3-106-122.

Статья поступила в редакцию 20.04.2023;
переработанный вариант – 01.08.2023;
принята к публикации 03.08.2023.

1. Введение

В настоящее время актуальной является задача создания надежной и анонимной системы дистанционного голосования. Большинство существующих систем голосования или не вполне отвечают требованиям анонимности, или являются закрытыми от исследования (в таком случае невозможно гарантировать их устойчивость к различным уязвимостям). Среди действующих систем голосования наиболее популярными и многофункциональными являются Google Forms [1] и Яндекс.Взгляд [2], но они могут применяться только для проведения открытых голосований. Некоторые сервисы, предлагающие пользователям создавать анонимные онлайн-опросы, не запрашивают имена голосующих напрямую, однако идентифицирующая

* Работа выполнена в рамках государственного задания № 071-03-2023-001 от 19.01.2023.

информация может быть раскрыта через различные метаданные, собираемые системой. Например, подобным недостатком обладает Survey Monkey [3] – популярный и востребованный сервис с широким функционалом. Одним из интересных проектов для построения системы анонимного голосования является ДЭГ (Дистанционное электронное голосование) [4], разрабатываемое компанией Ростелеком. В 2021–2022 г. в России проводились экспериментальные выборы с использованием данной системы. Она частично базируется на криптографических методах, например, подписание бюллетеней происходит с помощью «слепой» подписи [5], а анонимность выбора голосующих основывается на применении гомоморфного шифрования к бюллетеням, последующего суммирования зашифрованных данных на сервере и расшифровки всего суммированного шифротекста сразу с помощью ключа. Однако многие исследователи отмечают, что данная система не вполне гарантирует анонимность голосующих и подвержена достаточно серьезным уязвимостям. При подобной схеме построения системы голосования, основанной на гомоморфном шифровании, администраторам системы известен порядок появления бюллетеней на сервере, а также IP-адреса, с которых эти бюллетени были отправлены. В таком случае администраторы могут исключать по одному бюллетеню из суммы зашифрованных бюллетеней, складывать оставшиеся и таким образом постепенно узнавать, как проголосовали пользователи в порядке очереди [6]. Подобный процесс является довольно трудоемким, однако позволяет выяснить, как проголосовали конкретные участники, что ставит под сомнение анонимность системы. К тому же исследования, описанные в статьях [7] и [8], выявили значительные незадокументированные возможности ДЭГ, которые также вызывают вопросы о прозрачности выборов и потенциально могут нарушать тайну голосования.

При разработке эффективной схемы для создания систем анонимного голосования необходимо учитывать следующее: с одной стороны, голос каждого пользователя должен оставаться тайным, а с другой стороны, каждый должен иметь возможность проверить, что его голос был правильно учтён при подведении итогов голосования. Эта задача может быть решена с помощью криптографических средств.

Криптографические алгоритмы позволяют формировать, подписывать и впоследствии хранить бюллетени в электронном голосования. Для передачи бюллетеней участников на сервер необходимо создание анонимного канала. Распространенные решения для построения анонимных каналов имеют недостатки, например, возможность утечки данных пользователей (как в VPN [9]) или определения IP-адреса предыдущего узла (как в TOR [10]). Автором была разработана схема для создания системы анонимного дистанционного голосования, базирующаяся на протоколе «слепой» электронной подписи. Также в рамках схемы был разработан алгоритм анонимной доставки бюллетеней на сервер [11].

В предложенной схеме имеется некоторый сервер, выдающий участникам голосования бюллетени, подписанные с помощью «слепой» электронной цифровой подписи (ЭЦП). Особенностью этой разновидности электронной цифровой подписи является то, что подписывающая сторона не может знать точно содержимое подписываемого документа, она лишь проверяет этот документ на соответствие определенным условиям (параметрам) и, если условия выполняются, подписывает. Бюллетень представляет собой закодированную информацию о выборе пользователя в двоичном формате.

Основные требования к разрабатываемой системе заключаются в анонимности голосования, то есть ни сервер, ни кто-либо из участников не должен узнать, как проголосовал конкретный пользователь, а также в легитимности и уникальности бюллетеней. В данном случае это означает, что каждый участник может получить только один бюллетень, который невозможно подделать или использовать повторно. Легитимность и уникальность бюллетеней обеспечиваются криптографическими алгоритмами, а анонимность участников – каналом передачи бюллетеней на сервер. Канал состоит из сервера и узлов участников голосования, связанных по схеме «каждый с каждым». Когда голосующий сделал свой выбор, он отправляет бюллетень на сервер с вероятностью p либо кому-то из остальных участников с вероятностью $1–p$. Другой узел, получивший чужой бюллетень, поступает аналогично [11].

Однако в представленной схеме могут возникнуть определенные уязвимости, которые рассмотрены в п. 2. В п. 3 предлагается усовершенствованная схема анонимного голосования, устраняющая эти уязвимости. В п. 4 приведены расчеты, позволяющие выбрать наилучшее значение p – вероятности отправки бюллетеня на сервер, обеспечивающее наибольшую анонимность голосующих с точки зрения сервера.

2. Описание возможных уязвимостей протокола анонимного голосования

Для исследования надежности представленной схемы были рассмотрены уязвимости, которые могут возникнуть при построении и эксплуатации системы. Для системы голосований потенциально существует 2 класса угроз: уязвимости, допускаемые при описании системы, и уязвимости, возникающие на этапе ее программной реализации. Среди угроз, которые могут возникнуть при описании системы, также можно выделить подклассы уязвимостей:

- **Уязвимости, возникающие при описании системы**
  - Уязвимости, не зависящие от алгоритмов, заложенных при разработке системы голосования
    - Создание сервером фальшивых бюллетеней
    - возможный сговор всех участников против одного
  - Уязвимости при формировании и подписывании бюллетеня
    - «Мультипликативное свойство» алгоритма RSA
    - Повторное использование бюллетеней в одном голосовании
    - Использование бюллетеня из одного голосования в другом
  - Возможный саботаж при отправке данных по анонимному каналу и способы его устранения
- **Угрозы, возникающие при программной реализации системы**

Большее внимание стоит обращать на уязвимости, которые могут возникать на этапе теоретической разработки протоколов для организации всей системы голосований, чтобы исключить их появление при построении системы.

Среди данного класса в первую очередь можно выделить уязвимости, не зависящие от алгоритмов, заложенных при разработке системы голосования. К ним относится, например, настройка сервера таким образом, что он начнет создавать фальшивые бюллетени. Однако эта ситуация легко может быть обнаружена любым участником голосования. Если в сформированном по окончании голосования блоке количество бюллетеней не совпадет с количеством подписей участников, то, очевидно, произошел «вброс» бюллетеней, и данный блок с результатами голосования будет считаться недействительным. Другая уязвимость, не зависящая от алгоритма голосования, – возможный сговор всех участников против одного. В таком случае сговорившиеся могут узнать его выбор. Понятно, что возникновение такой ситуации всецело зависит от человеческого фактора.

К основным уязвимостям, которые необходимо исключить при описании системы голосования, относятся угрозы, приводящие к взлому криптографических алгоритмов. Например, «мультипликативное свойство» [12] алгоритма RSA, которое может позволить пользователю-злоумышленнику из двух легальных бюллетеней, подписанных сервером, сфабриковать новый нелегальный. Другая угроза заключается в возможности повторного использования бюллетеня, что приведет к «вбросу» бюллетеней. Еще одна угроза может возникнуть, когда пользователь берет легальный бюллетень из одного голосования и использует его в другом. В этом случае, как и в предыдущем, существует опасность получить неверный результат голосования. Все эти криптографические уязвимости были устранены в протоколе анонимного дистанционного голосования (версии 1), описанном автором в [11].

Однако также существует достаточно серьезная уязвимость, связанная с отправкой данных по анонимному каналу связи. Она заключается в возможности саботажа пользователем отправки своего или чужого бюллетеня далее по каналу. При этом невозможно отследить, какой именно пользователь саботировал процесс голосования. Для устранения этой уязвимости была создана новая версия протокола голосования, которая описана в п. 3.

Второй класс уязвимостей может возникнуть при практической реализации системы. Во-первых, она может быть разработана в любой среде с использованием любого программного средства, что вызывает необходимость внимательно учитывать особенности используемого языка программирования. Во-вторых, разработчик должен учитывать особенности, связанные с построением архитектуры баз данных, канала передачи данных, а также организации интерфейса, в частности, разделять интерфейс рядовых участников и администратора системы.

В настоящей работе предлагается устранение проблемы саботажа пользователем-злоумышленником отправки своего или чужого бюллетеня далее по каналу с помощью модифицированного протокола анонимного дистанционного голосования.

3. Протокол анонимного дистанционного голосования (версия 2)

Рассмотрим описание новой версии протокола голосования, представленного ранее в [11]. Пусть в системе голосований имеется сервер, который подписывает бюллетени при помощи ЭЦП. Описание процесса голосования можно разделить на четыре этапа, содержащих следующие шаги.

Этап 1. Формирование и подписывание бюллетеня

Шаг 1. В начале каждого голосования сервер генерирует общие данные согласно системе RSA [13]:

Выбираются случайные большие числа P (1024 бит) и Q (1024 бит), на их основе создаются числа

$$N = PQ \text{ и } \phi(N) = (P-1)(Q-1). \quad (1)$$

Затем выбирается случайное число $0 < D < \phi(N)$, взаимно простое с $\phi(N)$. Далее вычисляется:

$$C = D^{-1} mod\, \phi. \quad (2)$$

Таким образом, получаются секретные параметры сервера: $P_s = P$, $Q_s = Q$, $C_s = C$ и открытые параметры сервера: $N_s = N$, $D_s = D$.

Шаг 2. Аналогичным образом в начале каждого голосования все клиенты генерируют свои секретные и открытые ключи по алгоритму RSA:

$$(P_{ki}, Q_{ki}, C_{ki}, N_{ki}, D_{ki}), i = \overline{1,m}, \quad (3)$$

где m – число пользователей в системе.

Далее каждый пользователь высылает на сервер свой логин и открытые ключи N_{ki}, D_{ki}, сервер сохраняет полученные значения в открытую базу данных. Эти открытые ключи

впоследствии (на шаге 6) будут использоваться для доказательства честности проведенного голосования.

Шаг 3. Затем происходит процесс формирования и подписывания бюллетеней пользователей с результатами голосования по следующему алгоритму:

Шаг 3.1. Пользователь (назовем его Алиса) собирается проголосовать. Тогда клиентская сторона системы берет открытые ключи сервера N_s , D_s и приступает к созданию данных для подписи. Она генерирует случайное число rnd размером 512 бит (для создания уникального номера бюллетеня, чтобы исключить возможность повторного использования бюллетеня в одном голосовании) и выполняет конкатенацию rnd с числом v, которое является закодированным решением пользователя (данное число также не должно превышать 512 бит, кроме самого решения туда может быть включена служебная информация о голосовании). В результате получаем некоторое число n:

$$n = rnd \,|\, v \ (1024\,бит). \tag{4}$$

После этого формируются данные для подписи полученного числа n. Для этого сначала вычисляется криптографическая хэш-функция от числа n:

$$h = hash(n). \tag{5}$$

Затем генерируется случайное число r (маскирующий множитель) в промежутке $[2, N-1]$ и находится число $\overline{h}$:

$$\overline{h} = h\, r^{D_s} \ (mod\ N), \tag{6}$$

оно отправляется по защищенному каналу на сервер.

Шаг 3.2. Сервер помечает, что выдал бюллетень Алисе и «вслепую», то есть не зная точно содержимое бюллетеня, вычисляет подпись числа $\overline{h}$, пришедшего от данного пользователя:

$$\overline{s} = \overline{h}^{C_s} mod\ N. \tag{7}$$

Полученное значение отсылается обратно Алисе.

Шаг 3.3. Алиса снимает маскирующий множитель r с подписи своего бюллетеня по формуле

$$s = \overline{s}\, r^{-1} mod\ N, \tag{8}$$

где r^{-1} – это инверсия числа *r* по модулю *N*.

Шаг 3.4. Подписанный бюллетень (уже без маскирующего множителя) $\langle n_i, s_i \rangle$ должен быть отправлен на сервер голосования по анонимному каналу связи, описываемому далее (на этапе 2).

Алгоритм «слепой» электронной подписи бюллетеней, описанный на этапе 1, используется во многих криптографических протоколах. Существуют «плохие» (имеющие уязвимости) и «хорошие» варианты построения систем голосования, базирующихся на «слепой» подписи, однако в данной работе описана сразу «хорошая» схема, исключающая недостатки и уязвимости «плохих» схем.

Алгоритм слепой подписи, описанный в данной работе, основан на протоколе RSA, но RSA может быть заменен любым другим алгоритмом цифровой подписи с открытым и закрытым ключом, например, любым из алгоритмов семейства DSA [14]. Однако в нашем случае может возникнуть угроза использования злоумышленником «мультипликативного свойства» алгоритма RSA. Суть мультипликативного свойства заключается в возможности сфабриковать фальшивую банкноту путем перемножения двух настоящих по модулю N. Для борьбы с этой уязвимостью алгоритма на шаге 3.1 сервером подписывается не просто основная часть бюллетеня с выбранным пользователем вариантом ответа на голосование и служебной информацией внутри, а значение хеш-функции, вычисляемое от n. Следует выбирать хеш-функцию, не обладающую мультипликативным свойством, чтобы создать новый бюллетень посредством перемножения двух старых стало невозможно. Вычисление значения хеш-функции от n нужно также для того, чтобы банк не мог узнать номер банкноты, которую он подписывает «вслепую», так как криптографическая хеш-функция является односторонней функцией (то есть функцией, которая вычисляется легко, а обратная к ней – сложно).

Маскирующий множитель r служит также для гарантии анонимности пользователя, так как без него серверу присылалось бы только значение хеш-функции h. В таком случае сервер смог бы на этапе создания бюллетеней запомнить, какому пользователю соответствует каждое из хеш-значений, и затем, когда клиенты на этапе 2 пришлют свои бюллетени по анонимному каналу, вычислять значения хеш-функций от присылаемых пользователями бюллетеней n и узнавать, как проголосовал каждый из участников. После того, как сервер подписывает «замаскированный» бюллетень, клиент снимает маскирующий множитель и снова отправляет на сервер подписанный бюллетень уже по анонимному каналу (рис. 1).

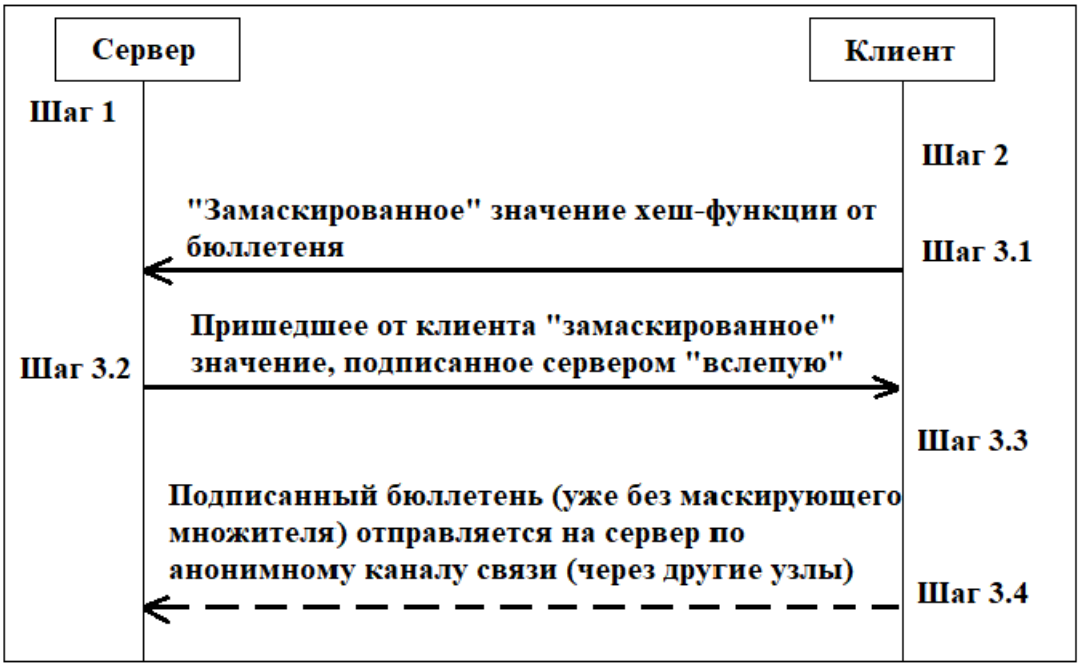

Рис. 1. Схема формирования и подписывания бюллетеней (этап 1)

Этап 2. Отправка бюллетеня по анонимному каналу связи на сервер

В предлагаемой системе отправка бюллетеней на сервер происходит по следующей схеме:

1. С вероятностью p бюллетень отправляется от пользователя на сервер, а с вероятностью $1–p$ бюллетень отправляется любому другому участнику голосования.

2. Участник, получивший чужой бюллетень по каналу связи, вычисляет значение хеш-функции от этого пришедшего бюллетеня, подписывает его своим секретным ключом и отправляет обратно приславшему.

Затем он с вероятностью p снова отправляет бюллетень на сервер или с вероятностью $1–p$ отправляет бюллетень любому другому участнику голосования. Пункты 1–2 повторяются до тех пор, пока бюллетень не окажется на сервере (рис. 2).

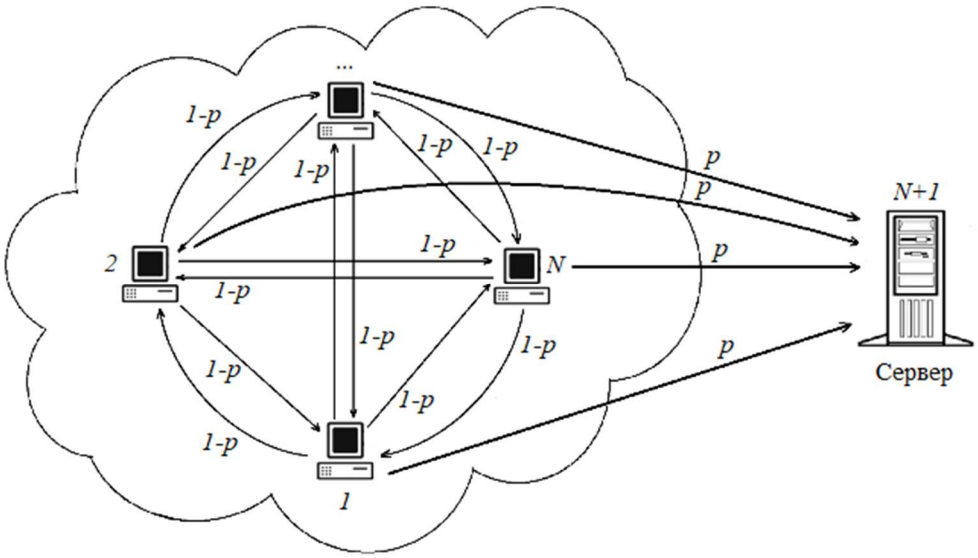

Рис. 2. Схема передачи бюллетеней в системе

Этап 3. Проверка корректности бюллетеней и отслеживание возможного саботажа процесса голосования

Шаг 4. Сервер проверяет корректность бюллетеня с помощью равенства:

$$hash(n_i) = s_i^{D_s} \bmod N \tag{9}$$

и в случае корректности учитывает голос и заносит информацию о бюллетене в открытую базу данных. Открытой она должна быть для того, чтобы любой участник мог убедиться, что его бюллетень был учтен.

Шаг 5. Когда время, выделенное на голосование, закончится и все отправленные пользователями бюллетени дойдут до сервера, сервер высылает всем клиентам список всех бюллетеней Z в случайном порядке.

Каждый пользователь, получив эту информацию, на своей стороне проверяет, что его бюллетень действительно находится в списке всех бюллетеней Z и вычисляет значение хеш-функции от этого списка:

$$HZ = hash(Z). \tag{10}$$

Затем пользователь подписывает своим секретным ключом хеш-значение от списка бюллетеней, т.е. вычисляет:

$$\overline{Z_i} = HZ^{C_{ki}} mod\ N_{ki}, \tag{11}$$

и высылает серверу полученное число $\overline{Z_i}$.

Если пользователь на данном этапе обнаружил, что его бюллетеня нет в общем списке бюллетеней Z, то он должен доказать, что его бюллетень был отправлен другому пользователю. В качестве доказательства предоставляется пришедший от пользователя, которому был переслан бюллетень, подписанный им хеш полученного бюллетеня. И так далее по цепочке до сервера. На данном этапе станет понятно, кто не отправил бюллетень дальше.

Этап 4. Формирование нового блока, хранящего результаты прошедшего голосования

Шаг 6. Сервер получает от всех пользователей числа $\overline{Z_i}$ (которые, по сути, являются подписями участников о том, что их бюллетени действительно находятся в списке всех бюллетеней Z). Затем сервер приступает к формированию блока, содержащего всю информацию о прошедшем голосовании. Блок имеет структуру вида:

$$\begin{aligned}&[\text{Хеш предыдущего блока}, Z,\\ &\left(\text{логин}_1\left(\text{или имя пользователя}_1\right), \overline{Z_1}\ , D_{k1}\right),\\ &\left(\text{логин}_2\left(\text{или имя пользователя}_2\right), \overline{Z_2}\ , D_{k2}\right),\\ &\qquad\qquad \ldots\\ &\left(\text{логин}_n\left(\text{или имя пользователя}_m\right), \overline{Z_m}\ , D_{km}\right)]\end{aligned} \tag{12}$$

Тогда любой пользователь в системе может проверить, действительны ли подписи всех остальных участников голосования, проверив равенство $hash(Z) = \overline{Z_1}^{D_{ki}} \bmod N_{ki}$. Таким образом можно удостовериться, что в блок не были добавлены лишние бюллетени. В результате голосования будут учитываться только голоса из бюллетеней, находящихся в сформированном блоке.

4. Определение степени анонимности сети

Максимально возможную анонимность канала передачи бюллетеней можно вычислить с помощью информационной энтропии – количественной меры неопределенности некоторой системы [15]. Для случайной величины x, принимающей n независимых случайных значений x_i с вероятностями p_i $i = (1,\ldots,n)$, информационная энтропия рассчитывается по формуле Шеннона:

$$H\left(x\right) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{13}$$

В предложенной системе голосования анонимность бюллетеня, пришедшего на сервер, зависит от вероятности p, с которой любой узел отправляет сообщение на сервер. Очевидно, что наименьшая анонимность отправителя бюллетеня с точки зрения сервера достигается при $p \to 1$. В таком случае с очень высокой вероятностью бюллетень проголосовавшего участника отправится сразу на сервер, не проходя через другие узлы, и значение неопределенности, выраженное с помощью энтропии, будет стремиться к 0 (тогда сервер может с большой вероятностью определить, как проголосовал каждый из участников).

Наибольшая анонимность отправителя бюллетеня с точки зрения сервера может быть достигнута в случае, если вероятность того, что автором сообщения был его отправитель на сервер, будет равна $1/N$, где N – количество клиентских узлов.

Можно рассчитать вероятность того, что автором сообщения был его отправитель на сервер, однако для разного количества клиентских узлов в системе формула для нахождения данной вероятности выражается по-разному.

4.1. Определение наилучшей вероятности отправки сообщения для минимального количества голосующих

Сначала рассмотрим самую простую конфигурацию сети из 2 клиентских узлов и 1 сервера (рис. 3) и рассчитаем значение p, при котором вероятность того, что автором бюллетеня был его отправитель, равна 0.5 (то есть оба пользователя равновероятно могли быть создателями бюллетеня). В таком случае энтропия принимает максимальное значение для данной конфигурации.

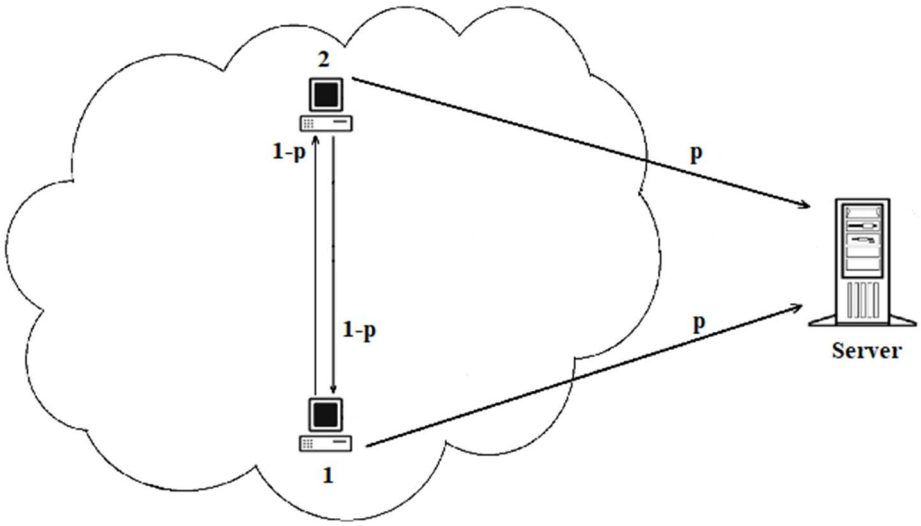

Рис. 3. Схема передачи бюллетеней в системе с 2 клиентскими узлами

Если бюллетень пришел на сервер от участника 1, то вероятность, что именно данный пользователь его сформировал, складывается из следующих вариантов: либо он сразу отправил его на сервер, либо бюллетень дошел от узла 1 до узла 2, вернулся обратно и был отправлен на сервер, при этом пересылаться между узлами 1 и 2 бюллетень может сколько угодно раз. Тогда данную вероятность можно описать формулой:

$$P(\text{автор бюллетеня = узел 1}) = p + p(1-p)^2 + p(1-p)^4 + p(1-p)^6 + \ldots + p(1-p)^{2n} = \sum_{n=0}^{\infty} p(1-p)^{2n}. \quad (14)$$

Из (14) видно, что вероятности несовместных событий, из которых складывается вероятность $P(\text{автор бюллетеня} - \text{узел 1})$, представляют собой геометрическую прогрессию: $b_1 = p$ – первый член прогрессии, $q = (1-p)^2$ – знаменатель прогрессии. Сумму первых n членов геометрической прогрессии в нашем случае можно рассчитать по формуле:

$$S_n = \frac{b_1(1-q^n)}{1-q} = \frac{p\left(1-\left((1-p)^2\right)^n\right)}{1-(1-p)^2} = \frac{p\left(1-(1-p)^{2n}\right)}{1-\left(1-2p+p^2\right)} = \frac{p\left(1-(1-p)^{2n}\right)}{1-1+2p-p^2} =$$

$$= \frac{p\left(1-(1-p)^{2n}\right)}{p(2-p)} = \frac{1-(1-p)^{2n}}{2-p}. \quad (15)$$

Чтобы аналитически вычислить значение p, необходимо решить уравнение:

$$\frac{1-(1-p)^{2n}}{2-p} = 0.5. \quad (16)$$

Сумма первых n членов геометрической прогрессии в формуле (16) приравнивается к 0.5, так как для сети с 2 голосующими наибольшая энтропия на сервере достигается при $P(\text{автор бюллетеня = узел 1}) = P(\text{автор бюллетеня = узел 2}) = 0.5$.

Значение энтропии в таком случае рассчитывается следующим образом:

$$H_{max} \text{ (при двух участниках)} = -\sum_{i=1}^{n} p_i \log_2 p_i = -(0.5 \cdot \log_2 0.5 + 0.5 \cdot \log_2 0.5) = 1. \quad (17)$$

В реальной ситуации неизвестно, сколько раз бюллетень будет перемещаться между узлами, пока не попадет на сервер, поэтому значение p, при котором достигается максимальная энтропия, зависит от количества этих пересылок n, и наоборот. Мы можем аналитически решить это уравнение, подставляя небольшие значения n = 1, 2, 3, …, 10, … (см. табл. 1). Очевидно, что при $n = 0$ бюллетень сразу попадает на сервер. Решая уравнение (16) для различных n, можно получить наилучшее значение вероятности p при разном количестве переходов между узлами. При таких наилучших значениях p энтропия на сервере всегда ≈ 1 (является наилучшей для сети, состоящей из 2 голосующих и сервера).

Таблица 1. Результаты аналитического решения уравнения (16)

n – количество проходов бюллетеня до узла 2 и обратно	p – вероятность отправки бюллетеня на сервер	Формула для расчета P (автор бюллетеня = узел 1)
0	0.5	P(а.б. = узел 1) = $p = 0.5$
1	≈ 0.352201	P(а.б. = узел 1) = $p + p(1-p)^2 \approx 0.5$
2	≈ 0.279587	P(а.б. = узел 1) = $p + p(1-p)^2 + p(1-p)^4 \approx 0.5$
3	≈ 0.234883	P(а.б. = узел 1) = $= p + p(1-p)^2 + p(1-p)^4 + p(1-p)^6 \approx 0.5$
…	…	…
10	≈ 0.120032	P(а.б. = узел 1) ≈ 0.5

При больших значениях n аналитическое решение уравнения становится трудоемким и не вполне целесообразным, так как данное уравнение удобнее решать программным путем. При каждом конкретном значении n формула (14) представляет собой монотонно возрастающую функцию [16], поэтому для решения уравнения (16) автором был программно реализован метод дихотомии [17]. При таком способе выбор значения n ограничен только вычислительной способностью компьютера.

Результаты, полученные с помощью имитационного (программного) моделирования, сведены в табл. 2. При этом энтропия всегда ≈ 1, P(автор бюллетеня = узел 1) ≈ 0.5.

Таблица 2. Результаты программного решения уравнения (16)

n – количество проходов бюллетеня до узла B и обратно	p – вероятность отправки бюллетеня на сервер
0	≈ 0.499938
1	≈ 0.352294
2	≈ 0.279541
3	≈ 0.234863
…	…
10	≈ 0.120117
…	…
100	≈ 0.022095
…	…
1000	≈ 0.003173
…	…
1000000	≈ 0.000244

Конечно, при $n = 1000000$ энтропия близка к 1, а вероятность p близка к 0. Однако в реальной системе такая ситуация недопустима в связи с огромными временными затратами. Для того, чтобы сервер не смог определить автора бюллетеня, достаточно всего нескольких пересылок между узлами. Очевидно, n и p – взаимозависимые величины, поэтому при настройке системы, состоящей из 2 узлов и 1 сервера, необходимо выбирать p в диапазоне [0.1, 0.25] (см. табл. 2).

4.2. Определение наилучшей вероятности отправки сообщения для трех и более голосующих

Теперь рассмотрим следующую конфигурацию сети: из 3 клиентских узлов и 1 сервера (рис. 4).

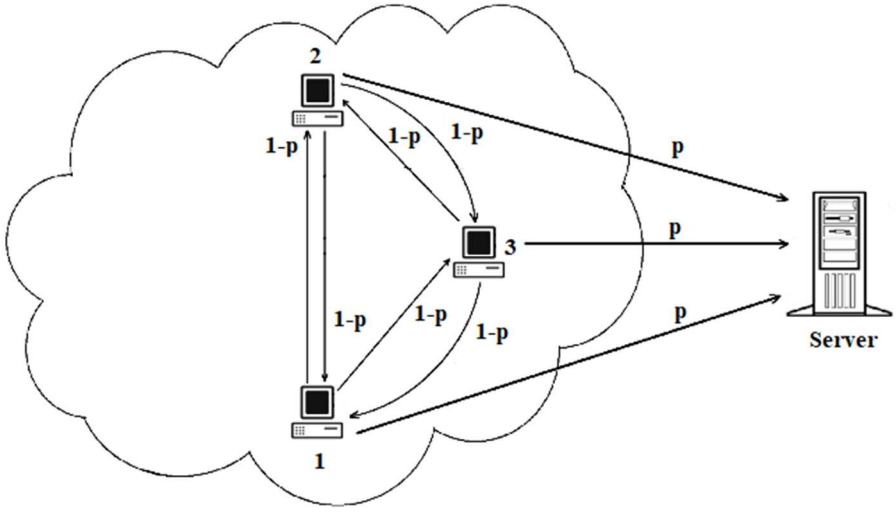

Рис. 4. Схема передачи бюллетеней в системе с 3 клиентскими узлами

В таком случае, если бюллетень пришел на сервер от участника 1, то вероятность, что именно данный пользователь его сформировал, складывается из следующих вариантов: либо он сразу отправил его на сервер, либо бюллетень пересылался несколько раз между другими участниками и в конце концов от участника 1 поступил на сервер. В данном случае определение вероятности $P(\text{автор бюллетеня} = \text{один из узлов})$ осложняется большим количеством

возможных переходов между узлами-участниками. Тогда данную вероятность можно описать формулой:

$$P\left(\text{автор бюллетеня = один из узлов}\right) = p + K_1\, p\left(\frac{1-p}{N-2}\right)^2 + K_2\, p\left(\frac{1-p}{N-2}\right)^3 + K_3\, p\left(\frac{1-p}{N-2}\right)^4 + \\ + K_4\, p\left(\frac{1-p}{N-2}\right)^5 + K_5\, p\left(\frac{1-p}{N-2}\right)^6 + K_6\, p\left(\frac{1-p}{N-2}\right)^7 + \ldots + K_j\, p\left(\frac{1-p}{N-2}\right)^{j+1}, \qquad (18)$$

где N – количество узлов в системе (клиенты + сервер), j – количество промежуточных узлов в цикле переходов бюллетеня, K_j – количество комбинаций возможных циклов с j промежуточными узлами.

$$K_j = \begin{Bmatrix} 2^j, & j = 1 \\ 2^j - K_{j-1}, & j \geq 2 \end{Bmatrix}. \qquad (19)$$

Например, для цикла с одним промежуточным узлом $K_1 = 2$, т.е. существуют две возможные комбинации прохождения бюллетеня через один узел и возвращения в исходный:

1 – 2 – 1
1 – 3 – 1.

Для цикла с двумя промежуточными узлами $K_2 = 2$:

1 – 2 – 3 – 1
1 – 3 – 2 – 1.

Для цикла с тремя промежуточными узлами $K_3 = 6$:

1 – 2 – 1 – 2 – 1
1 – 2 – 1 – 3 – 1
1 – 2 – 3 – 2 – 1
1 – 3 – 1 – 3 – 1
1 – 3 – 1 – 2 – 1
1 – 3 – 2 – 3 – 1.

Для сети с тремя голосующими наибольшая энтропия на сервере достигается при P(автор бюллетеня = узел 1) = P(автор бюллетеня = узел 2) = P(автор бюллетеня = узел 3) = 0.333. Тогда $P\left(\textit{автор бюллетеня = один из узлов}\right)$ приравнивается к 0.333:

$$p + K_1\, p\left(\frac{1-p}{2}\right)^2 + K_2\, p\left(\frac{1-p}{2}\right)^3 + K_3\, p\left(\frac{1-p}{2}\right)^4 + + K_4\, p\left(\frac{1-p}{2}\right)^5 + K_5\, p\left(\frac{1-p}{2}\right)^6 + \\ + K_6\, p\left(\frac{1-p}{2}\right)^7 + \ldots + K_j\, p\left(\frac{1-p}{2}\right)^{j+1} = 0.333. \qquad (20)$$

При трех и более пользовательских узлах нецелесообразно определять значение p аналитически, для решения уравнения (20) удобнее воспользоваться методом дихотомии,

программно реализованным автором. Результаты программного расчета приведены в табл. 3. При этом энтропия всегда ≈ 1.585:

$$H_{max} \text{ (при трех участниках)} = -\sum_{i=1}^{n} p_i \log_2 p_i = -3 \cdot (0.333 \cdot \log_2 0.333) \approx 1.585. \qquad (21)$$

Таблица 3. Результаты программного решения уравнения (20)

j – количество переходов бюллетеня между узлами участников голосования	p – вероятность отправки бюллетеня на сервер
0	≈ 0.333333
1	≈ 0.333251
2	≈ 0.261719
3	≈ 0.237549
…	…
7	≈ 0.165771
…	…
10	≈ 0.138306
…	…
50	≈ 0.050781
…	…
100	≈ 0.030762
…	…
1000	≈ 0.004883

В построенной системе голосований вероятность отправки бюллетеня сразу на сервер и количество проходов бюллетеня между узлами участников – взаимозависимые величины. Очевидно, что при малых значениях p потребуется большое количество пересылок бюллетеня, прежде чем он дойдет до сервера, что слишком затратно по времени. Для того, чтобы сервер не смог определить автора бюллетеня, достаточно всего нескольких пересылок между узлами. Из табл. 3 видно, что при 3 пользовательских узлах нужно выбирать p в диапазоне [0.05; 0.2].

Таким образом, при изменении количества участников голосования ($m > 2$) в системе должен происходить перерасчет величины p – вероятности отправки бюллетеня сразу на сервер. Для этого необходимо решить уравнение, в котором значение вероятности $P(\text{автор бюллетеня = один из узлов})$, определяемое по формуле (18), приравнивается к $1/m$, где m – количество участников голосования. При этом значение K_j вычисляется по формуле:

$$K_j = \begin{Bmatrix} (m-1)^j, & j = 1 \\ (m-1)^j - K_{j-1}, & j \geq 2 \end{Bmatrix}. \qquad (22)$$

Неопределенность же отправителя сообщения в узлах сети окажется максимальной независимо от выбранного значения p, потому что вероятности получения сообщения от любого из узлов сети одинаковы и равны $(1-p)/(m-1)$, где m – количество голосующих.

5. Заключение

В данной работе рассмотрены возможные угрозы и уязвимости протоколов, на которых базируется схема построения системы анонимного голосования. Противодействие почти всем уязвимостям предусмотрено в алгоритме создания анонимного дистанционного голосования [11]. Однако в ходе исследования была выявлена возможность саботажа злоумышленником дальнейшей пересылки бюллетеня. Для исключения подобных ситуаций алгоритм создания анонимного дистанционного голосования был усовершенствован. Новая версия алгоритма гарантирует невозможность подделки бюллетеней или их повторного использования, а также позволяет отслеживать попытки саботажа пересылки своих или чужих бюллетеней голосующими.

Также проведено исследование для определения максимально достижимой степени анонимности голосующих, которая может быть выражена с помощью информационной энтропии. Получены формулы для определения оптимального значения величины p – вероятности отправки бюллетеня сразу на сервер при разном количестве пользователей. Показано, что для пересылки данных по анонимному каналу необходимо вычислять p при заданном количестве голосующих, чтобы достичь максимально возможного значения энтропии. В этом случае достигается максимальная анонимность голосов пользователей с точки зрения сервера.

Литература

1. Google сервис для создания форм [Электронный ресурс]. URL: https://docs.google.com/forms (дата обращения: 01.08.2023).
2. Yandex сервис для создания опросов [Электронный ресурс]. URL: https://surveys.yandex.ru (дата обращения: 01.08.2023).
3. SurveyMonkey сервис для создания опросов [Электронный ресурс]. URL: https://www.surveymonkey.com (дата обращения: 01.08.2023).
4. Портал Дистанционного электронного голосования [Электронный ресурс]. URL: https://vybory.gov.ru/ (дата обращения: 01.08.2023).
5. *Chaum D., Rivest R. L., and Sherman A. T.* Blind Signatures for Untraceable Payments // Proc. Advances in Cryptology, Plenum (Springer-Verlag), New York, 1982. P. 199–203.
6. TAdviser - портал выбора технологий и поставщиков. Статья "Как устроено дистанционное электронное голосование в России, и какие у него перспективы". Глава "Тайна голосования" [Электронный ресурс]. URL: https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%9A%D0%B0%D0%BA_%D1%83%D1%81%D1%82%D1%80%D0%BE%D0%B5%D0%BD%D0%BE_%D0%B4%D0%B8%D1%81%D1%82%D0%B0%D0%BD%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D0%B3%D0%BE%D0%BB%D0%BE%D1%81%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%B2_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8,_%D0%B8_%D0%BA%D0%B0%D0%BA%D0%B8%D0%B5_%D1%83_%D0%BD%D0%B5%D0%B3%D0%BE_%D0%BF%D0%B5%D1%80%D1%81%D0%BF%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D1%8B#.D0.A2.D0.B0.D0.B9.D0.BD.D0.B0_.D0.B3.D0.BE.D0.BB.D0.BE.D1.81.D0.BE.D0.B2.D0.B0.D0.BD.D0.B8.D1.8F/ (дата обращения: 01.08.2023).
7. Хабр — сообщество IT-специалистов. Статья "Что же не так с ДЭГ в Москве?" [Электронный ресурс]. URL: https://habr.com/ru/post/579350/ (дата обращения: 01.08.2023).
8. Хабр — сообщество IT-специалистов. Статья "Что же не так с любыми электронными голосованиями?" [Электронный ресурс]. URL: https://habr.com/ru/post/579968/ (дата обращения: 01.08.2023).
9. Ferguson P., Huston G. What is a VPN? 1998: 22 [Электронный ресурс]. URL: https://www.potaroo.net/papers/vpn.pdf (дата обращения: 01.08.2023).

10. Dingledine R., Mathewson N., Syverson P. Tor: The second-generation onion router. Naval Research Lab Washington DC. 2004: 17 [Электронный ресурс]. URL: http://sharif.edu/~kharrazi/courses/40817-941/reading/tor-design.pdf (дата обращения: 01.08.2023).
11. *Dyachkova I., Rakitskiy A.* Anonymous remote voting system // Proc. International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk; 2019. P. 0850-0852.
12. *Рябко Б. Я., Фионов А. Н., Шокин Ю. И.* Криптография и стеганография в информационных технологиях. Новосибирск: Наука, 2015. 240 с.
13. *Rivest R., Shamir A., Adleman L.* A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM — NYC, USA: ACM. 1978. № 21 (2). P. 120–126. DOI:10.1145/359340.359342.
14. Kerry C.F., Gallagher P.D. Digital signature standard (DSS). FIPS PUB; 2013: 186-4 [Электронный ресурс]. URL: http://people.csail.mit.edu/alinush/6.857-spring-2015/papers/dsa.pdf (дата обращения: 01.08.2023).
15. Shannon C.E. A Mathematical Theory of Communication. Bell System Technical Journal. 1948; 27(3): 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338 [Электронный ресурс]. URL: https://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1948.tb01338.x (дата обращения: 01.08.2023).
16. *Ильин В. А., Садовничий В. А., Сендов Бл. Х.* Глава 4. Непрерывность функции // Математический анализ / Под ред. А. Н. Тихонова. 3-е изд., перераб. и доп. М.: Проспект, 2006. Т. 1. С. 146. 672 с.
17. *Левитин А. В.* Глава 11. Преодоление ограничений: Метод деления пополам // Алгоритмы. Введение в разработку и анализ. М.: Вильямс, 2006. С. 476—480. 576 с.

Дьячкова Ирина Сергеевна
старший преподаватель кафедры прикладной математики и кибернетики, Сибирский государственный университет телекоммуникаций и информатики (СибГУТИ, 630102, Новосибирск, ул. Кирова, д. 86), e-mail: dyach199@gmail.com, ORCID ID: 0000-0003-3696-1762.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

Development of a Secure Remote Voting System and Research on the Characteristics of its Anonymity

Irina S. Dyachkova

Siberian State University of Telecommunications and Information Science (SibSUTIS)

Abstract: Currently, the task of creating a reliable and anonymous remote voting system is relevant. Existing voting systems either do not fully meet the requirements of anonymity, or are closed from research. A scheme was developed for creating an anonymous remote voting system based on the blind electronic signature protocol and the algorithm for creating an anonymous channel. This paper describes possible threats and vulnerabilities of the protocols on which the scheme for constructing an anonymous voting system is based including the threat of a hacker's sabotage sending his own or other people's ballots over a communication channel. The presented optimized version of the algorithm for creating anonymous remote voting system guarantees the impossibility of fake ballots or their reuse and also excludes possible attempts to sabotage the forwarding of ballots by voters. Formulas for calculating the optimal value of p - the probability of sending

a ballot directly to the server is defined. It is shown that the value of p must be calculated depending on the number of voters in order to achieve the maximum anonymity of the participants' votes.

Keywords: anonymous voting, cryptography, anonymous data transmission, information entropy.

For citation: Dyachkova I. S. Development of a secure remote voting system and research on the characteristics of its anonymity (in Russian). *Vestnik SibGUTI*, 2023, vol. 17, no. 3, pp. 106-122. https://doi.org/10.55648/1998-6920-2023-17-3-106-122.

The article was submitted: 20.04.2023;
revised version: 01.08.2023;
accepted for publication 03.08.2023.

References

1. Google service for creating forms, available at: https://docs.google.com/forms (accessed 01.08.2023).
2. Yandex service for creating polls, available at: https://surveys.yandex.ru (accessed 01.08.2023).
3. SurveyMonkey service for creating polls, available at: https://www.surveymonkey.com (accessed 01.08.2023).
4. Remote Electronic Voting Portal, available at: https://vybory.gov.ru/ (accessed 01.08.2023).
5. Chaum D. Blind Signatures for Untraceable Payments. Chaum, D., Rivest R.L. and Sherman, A.T., Eds., Advances in Cryptology Proceedings of Crypto 82, *Plenum (Springer-Verlag)*, New York, p. 199-203.
6. TAdviser - a portal for choosing technologies and suppliers. *Kak ustroyeno distantsionnoye elektronnoye golosovaniye v Rossii, i kakiye u nego perspektivy. Tayna golosovaniya* [How remote electronic voting works in Russia, and what are its prospects. Secret of voting], available at: https://www.tadviser.ru/index.php/%D0%A1%D1%82%D0%B0%D1%82%D1%8C%D1%8F:%D0%9A%D0%B0%D0%BA_%D1%83%D1%81%D1%82%D1%80%D0%BE%D0%B5%D0%BD%D0%BE_%D0%B4%D0%B8%D1%81%D1%82%D0%B0%D0%BD%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D0%B3%D0%BE%D0%BB%D0%BE%D1%81%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%B2_%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B8,_%D0%B8_%D0%BA%D0%B0%D0%BA%D0%B8%D0%B5_%D1%83_%D0%BD%D0%B5%D0%B3%D0%BE_%D0%BF%D0%B5%D1%80%D1%81%D0%BF%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D1%8B#.D0.A2.D0.B0.D0.B9.D0.BD.D0.B0_.D0.B3.D0.BE.D0.BB.D0.BE.D1.81.D0.BE.D0.B2.D0.B0.D0.BD.D0.B8.D1.8F/ (accessed 01.08.2023).
7. Habr - a community of IT specialists. *Chto zhe ne tak s DEG v Moskve?* [What is wrong with the Remote Anonymous Voting in Moscow?], available at: https://habr.com/ru/post/579350/ (accessed 01.08.2023).
8. Habr - a community of IT specialists. *Chto zhe ne tak s lyubymi elektronnymi golosovaniyami?* [What is wrong with any electronic voting?], available at: https://habr.com/ru/post/579968/ (accessed 01.08.2023).
9. Ferguson P., Huston G. What is a VPN? 1998, 22 p., available at: https://www.potaroo.net/papers/vpn.pdf (accessed 01.08.2023).
10. Dingledine R., Mathewson N., Syverson P. Tor: The second-generation onion router. Naval Research Lab Washington DC, 2004. 17 p., available at: https://sharif.edu/~kharrazi/courses/40817-941/reading/tor-design.pdf (accessed 01.08.2023).
11. Dyachkova I., Rakitskiy A. Anonymous remote voting system. *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, Novosibirsk, 2019, pp. 0850-0852.
12. Ryabko B. Ya., Fionov A. N., Shokin Yu. I. *Kriptografiya i steganografiya v informatsionnykh tekhnologiyakh* [Cryptography and steganography in information technologies]. Novosibirsk, Nauka, 2015. 240 p.
13. Rivest R., Shamir A., Adleman L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM — NYC, USA: ACM*, 1978, 21(2), pp. 120-126. DOI:10.1145/359340.359342.

14. Kerry C. F., Gallagher P. D. Digital signature standard (DSS). FIPS PUB 186-4, 2013, available at: http://people.csail.mit.edu/alinush/6.857-spring-2015/papers/dsa.pdf (accessed 01.08.2023).
15. Shannon C. E. A Mathematical Theory of Communication. *Bell System Technical Journal*, 1948, 27(3), pp. 379-423, available at: https://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1948.tb01338.x (accessed 01.08.2023). DOI: 10.1002/j.1538-7305.1948.tb01338.
16. Ilyin V. A., Sadovnichiy V. A., Sendov Bl. Kh. *Matematicheskiy analiz. Tom 1* [Mathematical Analysis. vol. 1]. 3rd ed. Ed. A. N. Tikhonov. Moscow, Prospect, 2006. 661 p.
17. Levitin A.V. *Algoritmy. Vvedeniye v razrabotku i analiz* [Algorithms. Introduction to development and analysis]. Moscow, Williams, 2006. 576 p.

Irina S. Dyachkova

Senior lecturer of the Department of Applied Mathematics and Cybernetics, Siberian State University of Telecommunications and Information Science (SibSUTIS, Russia, 630102, Novosibirsk, Kirov St. 86), e-mail: dyach199@gmail.com, ORCID ID: 0000-0003-3696-1762.