DOI 10.37882/2223–2966.2021.03.38

# ЗАЩИТА ИНФОРМАЦИИ В ОБЛАЧНЫХ ТЕХНОЛОГИЯХ МЕТОДОМ РАССЕЧЕНИЯ-РАЗНЕСЕНИЯ

## PROTECTION OF INFORMATION IN CLOUD TECHONOLOGIES THE DISSECCTION-SEPARATION METHOD

***G. Shurkhovetsky***

*Summary.* Cloud technologies are now closely integrated into our lives in all its aspects. Users of cloud technologies are often not only individuals and legal entities, but also various government agencies, so that the security of these online services has the highest priority, including in matters of national security. But until now, the providers of these services can not provide adequate protection of user data, especially since by sending personal information to the "cloud", the user no longer controls it. An important role in this issue is played by the user's ability to protect their information on their side before sending it to various online services. An important tool for protecting information for the user will be such a tool that will provide reliable protection, taking into account the existence of several unrelated data warehouses. For example, the split-split method. On the Internet, this method has not yet been considered from the point of view of the reliability of its use, there was no proper verification of its resistance to attacks. This is the question that this article will be devoted to.

*Keywords:* information security, cloud storage, dissection-separation method.

***Шурховецкий Георгий Николаевич***
*Аспирант, ФГБОУ ВО «Иркутский государственный университет путей сообщения»*
*gshn5@yandex.ru*

*Аннотация.* Облачные технологии на сегодняшний день тесно вошли в нашу жизни во всех её аспектах. Пользователями облачных технологий зачастую выступают не только физические и юридические лица, но и различные государственные структуры, благодаря чему защищённость данных онлайн-сервисов имеет высший приоритет в том числе и в вопросах национальной безопасности. Но до сих пор провайдеры этих услуг не могут обеспечить достойную защиту данных пользователя, тем более что, отправив в «облако» личную информацию, пользователь её больше не контролирует. Важную роль в этом вопросе играет умение пользователя защищать свою информацию на своей стороне перед отправкой на различные онлайн-сервисы. Важным инструментом защиты информации для пользователя будет такой инструмент, который позволит обеспечить надёжную защиту с учётом существования нескольких не связанных между собой хранилищ данных. Например, метод рассечения-разнесения. В литературе данный метод до сих пор не был рассмотрен с позиций надёжности его использования, не было должной проверки его стойкости к атакам. Именно этому вопросу и будет посвящена данная статья.

*Ключевые слова:* защита информации, облачные хранилища, метод рассечения-разнесения.

На сегодняшний день всё большее распространение получают облачные технологии хранения данных, при этом при их использовании потребителю нельзя передавать или хранить информацию как личного, так и корпоративного характера. Потому что как только информация попадает в «облако», её уже пользователь не контролирует, она находится полностью в пользовании провайдеров, поставщиков данных услуг. То есть любая конфиденциальная информация (в том числе — касающаяся персональных данных) вообще не должна туда попадать, поскольку это — прямое нарушение Федерального закона 152 «О персональных данных». Что, безусловно, повышает риски для пользователей данных услуг. Отсюда возникает большой спрос на поиск эффективных методов защиты информации при её передаче и хранении. Как правило, при этом ограничиваются классическими криптографическими методами в виде шифрования, с последующим размещением защищаемой информации в единственном хранилище данных, при этом в остальных местах хранения могут храниться резервные копии. Однако этим возможные способы защиты не исчерпываются. Очевидно, что облачные технологии позволяют использовать одновременно несколько мест хранения, что расширяет арсенал возможных методов защиты информации. Большой интерес представляет разбиение информации на отдельные части и хранение этих, внешне бессмысленных частей, в разных географических точках. При этом информацию, в принципе, можно даже не зашифровывать в классическом смысле этого слова. То есть, наличие нескольких мест хранения привносит новые возможности защиты информации, один из которых — метод рассечения-разнесения — исследуется в данной статье. Интерес к данному методу обусловлен его алгоритмической простотой и естественностью в условиях массового распространения облачных технологий, однако его стойкость к атакам до сих пор подробно не рассматривалась [7]. Именно этому вопросу посвящена данная работа.

Таблица 1. Пример рассечения-разнесения текста

| 7 | 4 | 8 | 3 | 9 | 10 | 1 | 5 | 2 | 6 | 7 | 4 | 8 | 3 | 9 | 10 | 1 | 5 | 2 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Н | а | ш | е | й | _ | м | о | л | о | д | ё | ж | и | _ | ж | е | л | а | ю |
| 7 | 4 | 8 | 3 | 9 | 10 | 1 | 5 | 2 | 6 | 7 | 4 | 8 | 3 | 9 | 10 | 1 | 5 | 2 | 6 |
| _ | н | е | _ | б | ы | т | ь | _ | н | ы | т | и | к | а | м | и | . | _ | Н |

И т.д. на протяжении всей длины сообщения

| 7 | 4 | 8 | 3 | 9 | 10 | 1 | 5 | 2 | 6 | 7 | 4 | 8 | 3 | 9 | 10 | 1 | 5 | 2 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| т | о | м | _ | — | _ | э | т | о | _ | с | а | м | о | е | _ | г | л | а | в |
| 7 | 4 | 8 | 3 | | | | | | | | | | | | | | | | |
| н | о | е | . | | | | | | | | | | | | | | | | |

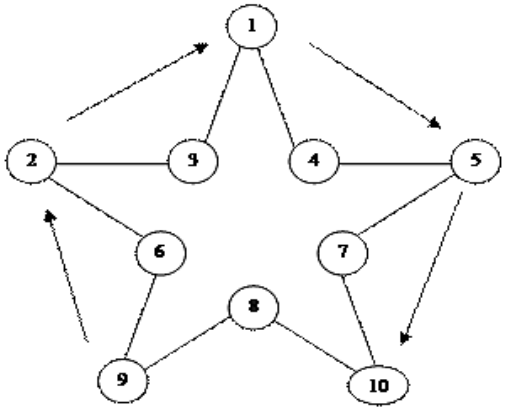



Рис. 1. Метод рассечения-разнесения в виде фигуры пятиконечной звезды

## Метод рассечения-разнесения

Из литературы известно, что существует такой метод криптографической защиты как рассечения-разнесения или рассечения (разнесения), который заключается в том, что исходное сообщение определённым образом разделяется на отдельные (равные между собой) блоки, которые затем разносятся в разные места хранения. Причём предполагается, что знание одного из потоков не позволяет восстановить исходного сообщения [4, 5].

В качестве примера возьмём текст:
«**Нашей_молодёжи_желаю_не_быть_нытиками._Нытиков_никогда_не_любил_—_и_к_себе_не_брал._Жизнь_должна_бить_ключом_и_за_любое_дело_нужно_браться_с_энтузиазмом._Я_всегда_оставался_ оптимистом_—_это_самое_главное.**» (А.Н. Ботян, Майор «Вихрь»).

Всего символов в сообщении 204. Зададим ключ рассечения-разнесения в виде числового набора {7, 4, 8, 3, 9, 10, 1, 5, 2, 6}. Ключ последовательно размещается вдоль текста по всей его длине, как показано в табл. 1. Каждое число — номер потока, в который помещается соответствующий символ. Легко видеть, что получившиеся потоки содержат внешне случайный и бессмысленный набор символов:

1. **мети_еиез_обусиваиэг**
2. **ла_ _илкбьи_ен_зеаиоа**
3. **еи_кка_е_жл_ оат_оо_о.**
4. **аёнттгиело_зебэмаяоао**
5. **оль.н_ _ _нбможяасвмтл**
6. **оюнНкю_р_ти_осм глс_в**
7. **Нд_ыыобсадь_д_ _одстсн**
8. **шжеиидлб.лкалрн._ _мме**
9. **й_бао_—_Жнюл_ту Ясп — е**
10. **_жымвн_ниачюньз_тт_ _**

Таким образом исходный текст заменяется десятью потоками, которые в сумме дают длину первоначального сообщения. Дальше полученные потоки разносятся в разные места хранения, причём всё это является обратимым процессом. Данная реализация метода рассечения-разнесения называется механическая (также существует смысловая) [2].

Для удобства запоминания ключ можно графически представить некоторой фигурой, например, звездой как на рис. 1. В неё вписывается в определённом порядке исходный текст (по 10 символов), а буквы зашиф-

рованного текста выписываются в заданной последовательности. К примеру, открытый текст записывается построчно, как изображено (пронумеровано) на рис. 1, а шифротекст считывается по кругу: 9–2–1–5–10–8–6–3–4–7 (показано стрелками). И т.д.

Таким образом, способ разбиения данных — это ключ метода, который состоит из непрерывной, случайно перемешанной последовательности натуральных чисел, начиная с единицы.

## Оценка криптостойкости

Стойкость данного метода рассмотрим по нескольким направлениям.

I. Ключ не известен, но при этом:
1. Известен один поток сообщения.
2. Известно несколько (более одного, но не все) потоков.
3. Известны все потоки.

II. Ключ известен и также:
1. Известен один поток сообщения.
2. Известно несколько (более одного, но не все) потоков.
3. Известны все потоки.

Очевидно, что это исчерпывает все возможные случаи. При этом наиболее вероятной выглядит ситуация I.1, наименее — ситуация II.3, когда у злоумышленника есть всё необходимое для восстановления сообщения. Последний случай тривиальный и его не рассматриваем.

Данное деление может выглядеть избыточным, однако остановимся на нём. Разберём ситуации последовательно.

**1. Ключ неизвестен и известен один поток (направление I.1)**

Если злоумышленнику известен один поток исходного сообщения, тогда ему необходимо для восстановления первоначального текста знать:

а) всю длину исходного текста;
б) в каком месте текста стоит конкретный символ известного потока;
в) подставлять символы в оставшиеся (пустые) места, чтобы восстановить сообщение, т.е. комбинировать.

Злоумышленник имеет один поток из общего их числа ($L$). Допустим, он знает, что защита информации была произведена методом рассечения-разнесения. В результате он предполагает, что исходная информация была рассечена на потоки и разнесена в места хранения каким-то образом.

Особенностью данного метода является то, что потоки, как правило, имеют равную длину, расхождения могут быть незначительны (либо +1 символ, если известный поток имеет $min$ длину, либо –1 — если $max$ длину, в примере выше, 4 потока длиной 21 символ, и 6 потоков — 20 символов). Если обозначить длину доступного злоумышленнику потока за $X_L$, тогда длину исходного сообщения можно выразить как $N=X_1+X_2+\\+X_3+\ldots+X_L$ или $X{\cdot}L$, где при достаточно большой длине все $X_i$ можно считать равными $X_L$. Также в зависимости от реализации метода деление на потоки исходного сообщения может происходить по определённому ключу, например, {7, 4, 8, 3, 9, 10, 1, 5, 2, 6} — номера десяти потоков, куда направляются символы в каждом блоке исходного текста (пример выше). Этот ключ, может быть постоянным, либо же каждый новый символ $m_z$ ($m$ – число символов в алфавите $D$, $m_z$ – это какой-то символ алфавита $D$ (исходный алфавит)) в каждом блоке исходного сообщения разносится постоянно в разные потоки в зависимости от нового ключа. Если обозначить ключ за $P=\{p_1,\ p_2,\ \ldots,\ p_t\}$, где $t$ — количество символов в ключе, причём $p_1,\ p_2,\ \ldots,\ p_t$ обозначает номер потока, куда пойдут символы каждого блока исходного сообщения после их рассечения. Общее количество вариантов возможных ключей равно [2]:

$$A_t=t! \tag{1}$$

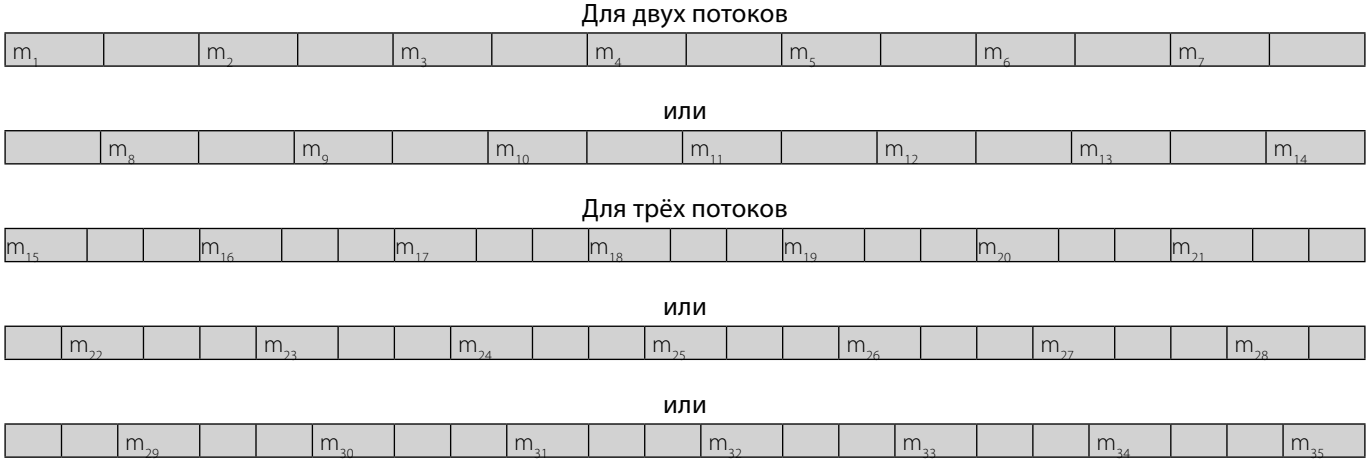
Как уже говорилось ранее, при постоянном ключе рассечения происходят равномерно (на потоки равной длины и с равной периодичностью рассечения). Если посмотреть на табл. 1, то можно видеть, что текст исходного сообщения последовательно рассекается на 10 частей, которые разносятся на 10 потоков, с периодом через каждые 10 символов. Более простой случай, разнесения на два или три потока, продемонстрирован на схеме 1.

Полагаем, что злоумышленник действует методом грубой силы [13–15]. Тогда при двух потоках (один из которых известен) количество возможных вариантов размещения символов в пустых позициях, соответствующих второму (отсутствующему) потоку, можно выразить величиной:

$$\bar{A}_m^{X_L} = 2 \cdot m^{X_L}, \tag{2}$$

где $m$ — это число символов в словаре (к примеру, $m$ = 54 для русского словаря, включая цифры и знаки препинания), $X_L$ — длина одного потока [1]. Здесь учтено

**Схема 1. Расположение символов известного злоумышленнику потока в тексте исходного сообщения, для двух или трёх потоков, может выглядеть, например, так (символы условные)**




(см. схему 1), что известные символы могут размещаться только в двух возможных позициях — в начале или в конце пары.

Общее же число комбинаций при $t$ возможных потоках будет рассчитываться как:

$$\bar{A}_m^{X_L} = t \cdot m^{X_L \cdot (t-1)}, \quad (3)$$

где $m^{X_L \cdot (t-1)}$ — количество вариантов заполнения пустых позиций для $t$-$l$ неизвестных потоков умноженное на число вариантов размещения букв известного сообщения, состоящего из $t$ потоков. Получится $t \cdot m^{X_L \cdot (t-1)}$. И то, если известно число потоков. Если не известно (а это так), ему придётся перебрать $t$ = 2, 3 и т.д. насколько хватит вычислительных возможностей и времени для взлома, пока информация не потеряет свою актуальность. Если он будет исходить из предположения, что потоков не более $M$, тогда число рассматриваемых вариантов будет:

$$\bar{A}_m^{X_L} == 2 \cdot m^{X_L} + 3 \cdot m^{2 \cdot X_L} + \cdots +$$

$$+ t \cdot m^{(t-1) \cdot X_L} + \cdots + + M \cdot m^{(M-1) \cdot X_L}.$$

Или, что тоже самое:

$$A = \sum_{t=2}^{M} t \cdot m^{(t-1) \cdot X_L} \quad (4)$$

По данному направлению криптоатака будет очень сложной. Хотя для коротких сообщений и небольшого числа потоков раскрыть сообщение всё-таки представляется возможным за короткое время. Например, в случае если исходное сообщение было поделено только на 2 потока с количеством символов не более 8, или 3–4 потока с количеством символов 3–4, то взломать можно за допустимое время для хорошего компьютера. Число возможных комбинаций будет составлять примерно от тысячи миллиардов до ста тысяч миллиардов (а если при переписке не изменялся ключ, тогда узнав его можно вскрыть и всю переписку), для сообщений длиной от 12 до 16 символов. Также если нам известен 1 поток с 4 символами, то все возможные комбинации для 3-х потоков (12 символов) можно рассчитать за допустимое время; количество вариантов здесь: $4 \cdot 3 \cdot 54^{4 \cdot 2} \approx 9 \cdot 10^{14}$. Только исходное сообщение крайне короткое, но злоумышленник может узнать благодаря такому короткому сообщению способ, по которому происходили рассечения-разнесения, тогда, в случае, если данный текст взят из переписки и сам способ на протяжении переписки не изменялся, можно будет понять на каких позициях стоят известные символы и возможно удастся вскрыть всю переписку или её часть.

**2. Ключ неизвестен и известно несколько (но не все) потоков (направление I.2)**

Когда злоумышленнику известно какое-то количество потоков ($l$). Больше он ничего не знает.

Злоумышленник предполагает, что всего потоков не более $M$, где $M > l$. Тогда для некоторого промежуточного значения $l < t \leq M$ символы известных потоков можно разместить

$$C_t^l = \frac{t!}{(t-l)! \cdot l!} \quad (5)$$

числом способов; $t$–$l$ остальных символов неизвестных потоков придётся подбирать. В результате количество вариантов для взлома грубой силой для $t-l$ потоков равно:

$$A = t \cdot m^{(t-l) \cdot X_L} \quad (6)$$

Если предположить, что потоков не более $M$, общее число возможных вариантов станет:

$$A = \sum_{t=l+1}^{M} \frac{t!}{(t-l)! \cdot l!} \cdot t \cdot m^{(t-l) \cdot X_L} \quad (7)$$

Данная формула (7) характерна для случаев, когда злоумышленнику известно ≤50% из всех потоков, а также, когда > 50% из всех потоков.

Множество $f\{X_1, X_2, \ldots, X_l\}$ потоков известное злоумышленнику, $l$ — общее количество известных потоков [9–12]. Можно предположить два варианта развития криптоатаки, когда злоумышленник знает большинство потоков исходного сообщения, например, 70–80%, либо же меньшинство 20–30%. Изначально можно скомбинировать известное количество элементов, предполагая, что ему досталось большинство потоков. Если это так, тогда имея порядка 70–80% потоков исходного сообщения, можно вскрыть исходное сообщение, потому что если рассечения происходили с постоянным ключом, тогда учитывая периодичность рассечения, нужно будет по формуле перестановки без повторений (8) [3]:

$$A_{X_l} = l!, \quad (8)$$

получится общее число возможных комбинаций известных потоков.

Скомбинировать последовательность потоков между собой, определив сколькими способами можно разместить $Xl$ различных потоков. Произведя выборку из вариантов, можно будет увидеть, что в некоторых комбинациях получится больше осмысленных слов, и именно такие варианты и будут подходить для поиска исходного сообщения. Предполагая, что злоумышленнику неизвестно не более 20–30% потоков. Возможно, выявится сразу же недостающий пробел и можно будет добавить ещё поток с символом пробела, при прочтении получившегося текста, при отсутствии некоторых символов можно воспользоваться антиципацией (антиципация — предвосхищение), для достраивания исходного сообщения [6].

В том случае, если злоумышленнику будет известно 20–30% исходного, тогда вскрытие оставшейся части исходного сообщения будет происходить по аналогии с первым направлением криптоанализа, когда злоумышленнику известен лишь один поток, с той лишь разницей, что известно будет $l$ потоков.

**3. Ключ известен и известен один поток (направление II.1)**

Полагаем, что злоумышленнику известен один поток исходного сообщения и способ, по которому был реализован метод рассечения-разнесения, то есть число потоков. Этот случай сводится к 1.1, только число вариантов для взлома будет определяться формулой (3), так как суммирование по возможному числу потоков здесь отсутствует.

4. Ключ известен и известно несколько (но не все) потоков (направление II.2).

В этом случае вскрытие сводится к подбору неизвестных символов в пустых местах для всех вариантов расстановки $l$ известных символов сообщения. Т.к. общее число потоков злоумышленнику известно (устанавливается по ключу), число вариантов подбора равно:

$$A = \frac{t!}{(t-l)! \cdot l!} \cdot t \cdot m^{(t-l) \cdot X_L} \quad (9)$$

где $t$ — число потоков.

В данном случае сразу же будет известна длина сообщения, расположение известных символов из известных потоков на своём месте в исходном сообщении. И если известное количество известных потоков будет составлять порядка 70%, тогда остальные символы в большинстве своём угадываются интуитивно или посредством антиципации. (Данное утверждение каждый читатель может проверить сам по примеру, указанному в начале статьи, убрав любые 1–3 потока, текст всё равно поддаётся прочтению, либо же сформулировать свой пример и опробовать на нём данное утверждение.) А если предположить, что речь идёт о графическом файле, то можно воспользоваться, например, восстановлением изображения с помощью самоорганизующихся карт Кохонена. Причём восстановление возможно даже при 12% известных пикселей [8]. Остальные символы могут быть разгаданы по формуле размещение с повторением (2). При данном условии взлом будет успешным. А если же количество известных потоков составит не более 20–30% от исходного их числа, тогда вскрытие оставшейся части исходного сообщения будет происходить по аналогии со 2 направлением криптоанализа, когда злоумышленнику известен один поток исходного сообщения и способ, по которому был реализован метод рассечения-разнесения. Но при этом длина исходного текста также неизвестна.

**5. Ключ неизвестен, но известны все потоки (направление II.3)**

Когда злоумышленник знает все потоки и осведомлён об этом, он знает их количество. Тогда вскрытие сообщение сводится к подбору порядка первых символов, чтобы получить осмысленную

последовательность, таким образом получив ключ. Можно предположить, что восстановление исходного сообщения для большого числа потоков не составит труда. Для малого числа, например, двух или трёх, а также нетекстовых файлов осмысленное сообщение подобрать труднее, но это можно сделать для относительно длинной последовательности символов. В любом случае число вариантов выбора равно:

$$A = x_M!, \quad (10)$$

где $x_M$ — первые символы $x_1, x_2, \ldots, x_M$ потоков $M$, что делает вскрытие простым делом.

## Другие решения

Перспективным вариантом решения для защиты информации в реалиях облачных технологий можно предложить посимвольное (побайтовое) разделение исходного сообщения на потоки разной длины. Например, гласные в один поток, согласные в другой, знаки препинания в третий и цифры в четвёртый. В таком случае даже знание одного потока, нескольких потоков или всех потребует взлома грубой силой (в первых 2-х случаях по формуле размещение с повторениями (2) и в 3-м случае перестановка без повторений (10)).

## Заключение

Из вышесказанного можно сделать вывод, что данный метод защиты информации позволяет подойти к защите без явного шифрования на основе одного только факта раздельного хранения информации с использованием разных мест хранения. Однако, как показано в работе, сам метод рассечения-разнесения в вышеописанной реализации бывает уязвим, что требует его дальнейшего изучения и развития.

ЛИТЕРАТУРА


1. Виленкин Н.Я. Комбинаторика. М.: Наука. Гл. ред. физ.-мат. лит., 1969.— 323 с.
2. Лыгин Е.А. Тайнопись. Практическое пособие по ручному шифрованию.— 4-е изд., доп.— Саратов: издательство «Новый ветер», 2010.— 206 с.
3. Райгородский А.М. Линейно-алгебраический метод в комбинаторике Электронное издание М.: МЦНМО, 2015~<144 с.
4. Партыка Т.Л., Попов И.И. Информационная безопасность. Учебное пособие для студентов учреждений среднего профессионального образования.— М.: ФОРУМ: ИНФРА-М, 2002.— 368 с.: ил.— (Серия «Профессиональное образование»).
5. Безбогов А.А., Яковлев А.В., Шамкин В.Н. Методы и средства защиты компьютерной информации: Учебное пособие.— Тамбов: Издательство ТГТУ, 2006.— 196.
6. Антиципация // Энциклопедический словарь Брокгауза и Ефрона: в 86 т. (82 т. и 4 доп.).— СПб., 1890–1907.
7. Шурховецкий Г.Н. Защита информации во внешних хранилищах данных методом рассечения-разнесения [Электронный ресурс] / Г.Н. Шурховецкий // Молодая наука Сибири: электрон. науч. журн.— 2020.— № 3(9).— Режим доступа: http://mnv.irgups.ru/toma/39–2020, свободный.— Загл. с экрана.— Яз. рус., англ. (дата обращения: 26.12.2020)
8. Восстановление изображений при помощи нейросетей. [Электронный ресурс].— Режим доступа: https://habr.com/ru/post/120473/, свободный (Дата обращения 26.10.2020).
9. Devlin, K. Aspects of Constructibility. Springer 1984.
10. Drake F.R. Set Theory. An Introduction to Large Cardinals. North Holland 1974.
11. Felgner U. Models of ZF-set theory. Springer 1971.
12. Jech Th. The Axiom of Choice. Springer 1993.
13. Ayesha Siddiqa and Sohail Ahmed, “Scalable Asymmetric Security Mechanism for Internet of Things” International Journal of Advanced Computer Science and Applications (IJACSA), 11(8), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110847
14. P Rajesh, Mansoor Alam, Mansour Tahernezhadi, T. Ravi Kumar and Vikram Phaneendra Rajesh, “Secure Communication across the Internet by Encrypting the Data using Cryptography and Image Steganography” International Journal of Advanced Computer Science and Applications (IJACSA), 11(10), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111057
15. Afnan Alsadhan, Asma Alotaibi, Lulu Altamran, Majd Almalki, Moneera Alfulaij and Tarfa Almoneef, “Manar: An Arabic Game-based Application Aimed for Teaching Cybersecurity using Image Processing” International Journal of Advanced Computer Science and Applications (IJACSA), 11(10), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111051




Журнал «Современная наука: актуальные проблемы теории и практики»