

Разработка методики восстановления данных на запоминающих устройствах

Василий С. Замолоцких^{1*}, Валентина Г. Сидоренко¹

¹Российский университет транспорта, Москва, Российская Федерация

*netaonh11@gmail.com



Василий С.
Замолоцких



Валентина Г.
Сидоренко

Резюме. Цель. Развитие цифровых технологий обуславливает необходимость в оцифровке данных и их последующем хранении на электронных носителях. Вне зависимости от способа хранения информации информация представляет собой ценность, а ее утрата может повлечь за собой ущерб. Существует ряд профилактических мер (применение аппаратной и логической избыточности разных видов), чтобы предотвратить эту утрату. В том случае если профилактические мероприятия в силу причин и обстоятельств не обеспечили защиту данных и доступ к ним был потерян, возникает потребность в их полном и своевременном восстановлении. Исходя из этого, возникает необходимость в создании алгоритма восстановления информации, который учтет аппаратные особенности современных накопителей, их логическое устройство, а также специфику хранимых на нем данных. **Методы.** Имеется два подхода к восстановлению информации: универсальный и персональный. Универсальный предусматривает использование минимального количества программ и средств, которые работают со всеми объектами. Персональный подход обуславливает использование большого количества программ и средств, которые работают с конкретными проблемами, связанными с утратой доступа к информации. За счет этого достигается быстрое действие и высокое качество восстановления по сравнению с универсальным подходом. Кроме того, персональные программы, как правило, дешевле универсальных. Использование универсальных средств восстановления информации не дает высокого результата в случае их применения для большого количества вариантов аварийных событий. Одной утилиты может быть недостаточно для решения всех проблем, вызванных инцидентами. Готового шаблона, следуя которому можно получить приемлемый результат, тоже не существует. Помимо использования персонального программного обеспечения существуют дополнительные альтернативы универсальному подходу – это программы для ручного восстановления данных и программно-аппаратные комплексы. В случае незначительных логических сбоев (повреждение главной загрузочной записи) применяется программное обеспечение для ручного восстановления информации. В случае критических аппаратных проблем с накопителем применяются программно-аппаратные комплексы. **Результаты.** Создана методика восстановления данных на запоминающих устройствах разных типов, включающая в себя универсальный и персональный подходы к восстановлению информации, использование программного обеспечения для ручного восстановления данных, а также применение программно-аппаратных комплексов. Данная методика позволяет восстанавливать данные популярных расширений с распространенных файловых систем и накопителей, предусмотрено взаимодействие с RAID-массивами всех уровней. Произведен отбор программ из восьми выделенных множеств методом анализа иерархий с приоритетом критерия результативности. Проведен ряд тестовых испытаний данной методики. Для проведения тестирования выполнена эмуляция инцидентов, связанных с утратой доступа к данным. Дана экономическая оценка стоимости устранения различных инцидентов с использованием разработанной методики. **Выводы.** На основе полученных результатов тестирования сформулированы выводы об эффективности персонального подхода к восстановлению информации.

Ключевые слова: накопитель информации, организация данных, метод анализа иерархий.

Для цитирования: Замолоцких В.С., Сидоренко В.С. Разработка методики восстановления данных на запоминающих устройствах // Надежность. 2022. №1. С. 56-62. <https://doi.org/10.21683/1729-2646-2022-22-1-56-62>

Поступила 21.01.2022 г. / После доработки 03.02.2022 г. / К печати 18.03.2022 г.

Введение

Данные имеют первостепенное значение, и в определенных случаях их стоимость превышает стоимость накопителей, на которых они хранятся. Цена восстановления также может быть выше стоимости накопителей. Поэтому очень важно грамотно оценить информацию и определить, нуждается ли она в восстановлении, выгодно ли это с экономической точки зрения. Данная оценка, а также причина утраты доступа к информации помогут выбрать метод восстановления информации на запоминающих устройствах, описанный в данной статье.

Существует несколько определений восстановления данных. Восстановление данных – последовательность действий, направленных на извлечение информации с запоминающего устройства в случае, когда эта информация не может быть прочитана обычным способом [1].

Восстановление данных – это последовательность действий, при котором осуществляется доступ к поврежденной или нечитаемой информации, и она переносится на другое устройство [2]. Их основное отличие друг от друга состоит в том, что во втором определении напрямую указывается на необходимость переноса восстановленных данных на другое (имеется ввиду – исправное) устройство.

1. Организация данных на накопителях

Создание методики восстановления данных предусматривает анализ причин потери информации. Для классификации последних необходимо учитывать то, каким образом данные организованы на накопителе. Организация данных рассматривается с двух сторон, аппаратной и логической.

Аппаратная организация данных зависит от типа накопителя, его физической составляющей. Для накопителя на жестких магнитных дисках (НЖМД) они заносятся на магнитные пластины, которые располагаются в гермоблоке диска [3]. У полупроводниковых накопителей данные хранятся на плавающих затворах, которые используются в одноименных транзисторах. В состав полупроводниковых накопителей входят внешний интерфейс (в случае периферийного подключения к компьютеру), контроллер и ячейки памяти. [1]. В оптических накопителях считывающий луч проходит сквозь поликарбонатную основу, попадает на информационный слой и отражается. [4, 5].

Аппаратная организация данных оказывает влияние на причины потери информации.

1.2. Логическая организация

Логическая организация универсальна и не зависит от типа накопителя. Она представляет собой иерархию: диск, раздел, каталог, файл [4]. В первом секторе диска записывается *MBR* (главная загрузочная запись), которая необходима для загрузки операционной системы. Строе-

ние раздела зависит от объема накопителя. В накопителях объемом свыше 2 ТБ используется *GUID* (технология уникального идентификатора). В структуре файла имеется основной заголовок, который хранит сигнатуру файла. Программы для восстановления файлов используют данную запись для обнаружения файлов на накопителях.

2. Причины потери информации

Причины потери информации представлены на рис. 1 [6]. Основываясь на информации об организации данных, можно выделить 3 группы причин потери информации: аппаратные, логические и человеческий фактор (рис. 2).



Рис. 1. Распределение причин потерь информации

3. Алгоритм восстановления информации на запоминающих устройствах

Основываясь на анализе физической и логической организации данных на накопителях, а также причин потери информации, авторы создали алгоритм восстановления данных на запоминающих устройствах (рис. 3). Он основан на использовании аппаратной и логической избыточности организации данных, а также использовании аналогов накопителей. Аппаратная избыточность реализуется путем копирования служебных данных, резервирования и виртуализации, а логическая – добавления сигнатуры в заголовок файла. Аналоги исправных накопителей удобно использовать для восстановления данных ручным способом.

Выбор направления восстановления зависит от степени повреждения служебной информации, включающей в себя данные файловых таблиц, в которых находятся адреса файлов. Если они сильно повреждены, то основной упор делается на анализ битовых последовательностей, иначе говоря, на сигнатурный поиск. Стоит отметить, что данный метод самый популярный и большинство программ используют именно его, что, впрочем, не отменяет использование ими же служебной информации для восстановления данных. Ручное восстановление осуществляется при помощи шестнадцатеричных и дисковых редакторов. В случае восстановления служебных данных удобно использовать дисковый редактор, т.к. он представляет накопитель в виде образа, разбитого на сектора. Сигнатурный поиск можно осуществлять, используя окно поиска, занося в него сигнатуру файла [4].



Рис. 2. Причины потери информации

4. Отбор программного обеспечения посредством метода анализа иерархий

Для того чтобы алгоритм был работоспособен, его необходимо оснастить программным обеспечением (ПО). Имеющееся на рынке ПО, которое может быть задействовано при реализации этапов разработанного алгоритма, было разделено на 8 групп. В каждой из групп был выполнен выбор программ методом анализа иерархий (МАИ) [7] с приоритетом критерия результативности (табл. 1).

МАИ позволяет осуществлять попарное сравнение объектов, а также принимать решение, опираясь на критерии, характерные для объектов. Выбор этого метода обусловлен следующими преимуществами [8]: возможность попарного сравнения критериев, что облегчает вычисления, т.к. нет необходимости «держать в поле зрения» все критерии, которых может быть много;

наличие вербально-числовой шкалы, которая упрощает оценивание, т.к. сопоставляет числа с удобными для восприятия логическими сравнительными конструкциями; встроенная в метод оценка качества эксперта (отношение согласованности).

Формулы (1)–(3) позволяют провести оценку объектов по характерным для них критериям.

$$\text{CPГЕОМ}_x = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}, \quad (1)$$

где CPГЕОМ_x – среднее геометрическое значение суждений x (по строке);

x – положительное число;

n – количество чисел x .

$$\text{СУММ}_a = \sum_{i=1}^n a_i, \quad (2)$$

где СУММ_a – результат суммы чисел a ;

a_i – число с порядковым номером i ;

n – количество чисел a .

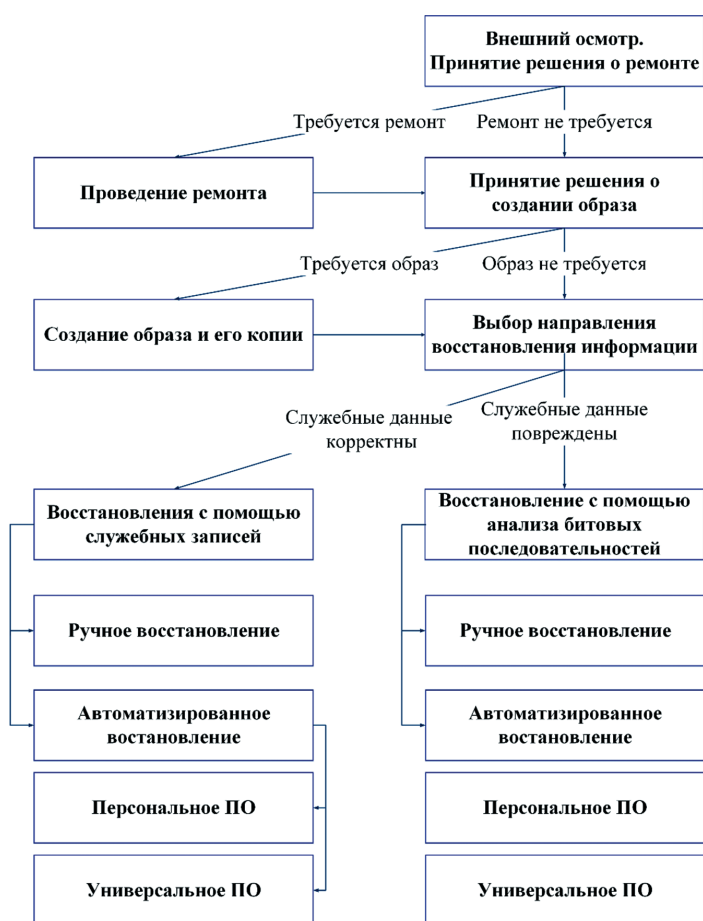


Рис. 3. Алгоритм восстановления информации

$$BEKTOP_x = \frac{CPГЕОМ_x}{СУММ_{CPГЕОМ_x}}, \quad (3)$$

где $BEKTOP_x$ – значение вектора приоритета критерия/ПО x ;

$CPГЕОМ_x$ – среднее геометрическое значение чисел x ;

$СУММ_{CPГЕОМ_x}$ – сумма средних геометрических значений x .

Итоговая оценка будет получена после применения формулы 4:

$$ГЛОБВЕКТОР_x = \sum_{i=1}^n BEKTOP_{критерий_i} \times BEKTOP_{ПО_i}, \quad (4)$$

где $ГЛОБВЕКТОР_x$ – значение глобального вектора приоритета ПО x ;

n – количество критериев;

$BEKTOP_{критерий_i}$ – значение вектор-приоритета критерия с порядковым номером i ;

$BEKTOP_{ПО_i}$ – значение вектор-приоритета ПО с порядковым номером i .

Группа формул (5)-(7) необходима для расчета отношения согласованности (ОС), которое определяет корректность суждений. Согласно рекомендациям, оно не должно превышать 10%, эта рекомендация была выполнена.

Табл. 1. Исходные данные и результаты работы МАИ

№ п/п	Тип ПО	Количество рассмотренных вариантов ПО	Победитель
1	Тестирование накопителей	5	R.tester
2	Создание побайтовых копий накопителей	5	Active Disk Image
3	Ручное восстановление данных	7	Acronis Disk Director
4	Универсальное	13	Hetman Partition Recovery
5	Восстановление информации с RAID-массивов	3	RS RAID Retrieve
6	Восстановление удаленных файлов	4	Active Undelete
7	Восстановление поврежденных файлов	12	RS File Repair, пакет ПО Recovery Toolbox
8	Персональное	25	Recuva, EasyRecovery

Табл. 2 . Оценка критериев МАИ

Критерий	Результативность	Универсальность по типу накопителей	Цена	Дополнительные возможности	Вектор-приоритет	Среднее геометрическое	ИС	ОС МАИ, %
Результативность	1,0	3,0	4,0	4,0	0,5285	2,6	0,03	3,52
Универсальность по типу накопителей	0,3	1,0	2,0	3,0	0,2388	1,2		
Цена	0,3	0,5	1,0	2,0	0,1420	0,7		
Дополнительные возможности	0,3	0,3	0,5	1,0	0,0907	0,5		
Сумма	1,8	4,8	7,5	10,0	1	5,0		

Табл. 3. Оценка ПО согласно критерию результативности

	Everest Ultimate Edition/AIDA 64	Victoria	MHDD	R.testер	HDDScan	Вектор-приоритет	Среднее геометрическое	ИС	ОС МАИ, %
Everest Ultimate Edition/AIDA 64	1,0	2,0	2,0	1,0	2,0	0,2857	1,5	0,00	0,00
Victoria	0,5	1,0	1,0	0,5	1,0	0,1429	0,8		
MHDD	0,5	1,0	1,0	0,5	1,0	0,1429	0,8		
R.testер	1,0	2,0	2,0	1,0	2,0	0,2857	1,5		
HDDScan	0,5	1,0	1,0	0,5	1,0	0,1429	0,8		
Сумма	3,5	7,0	7,0	3,5	7,0	1,0000	5,3		

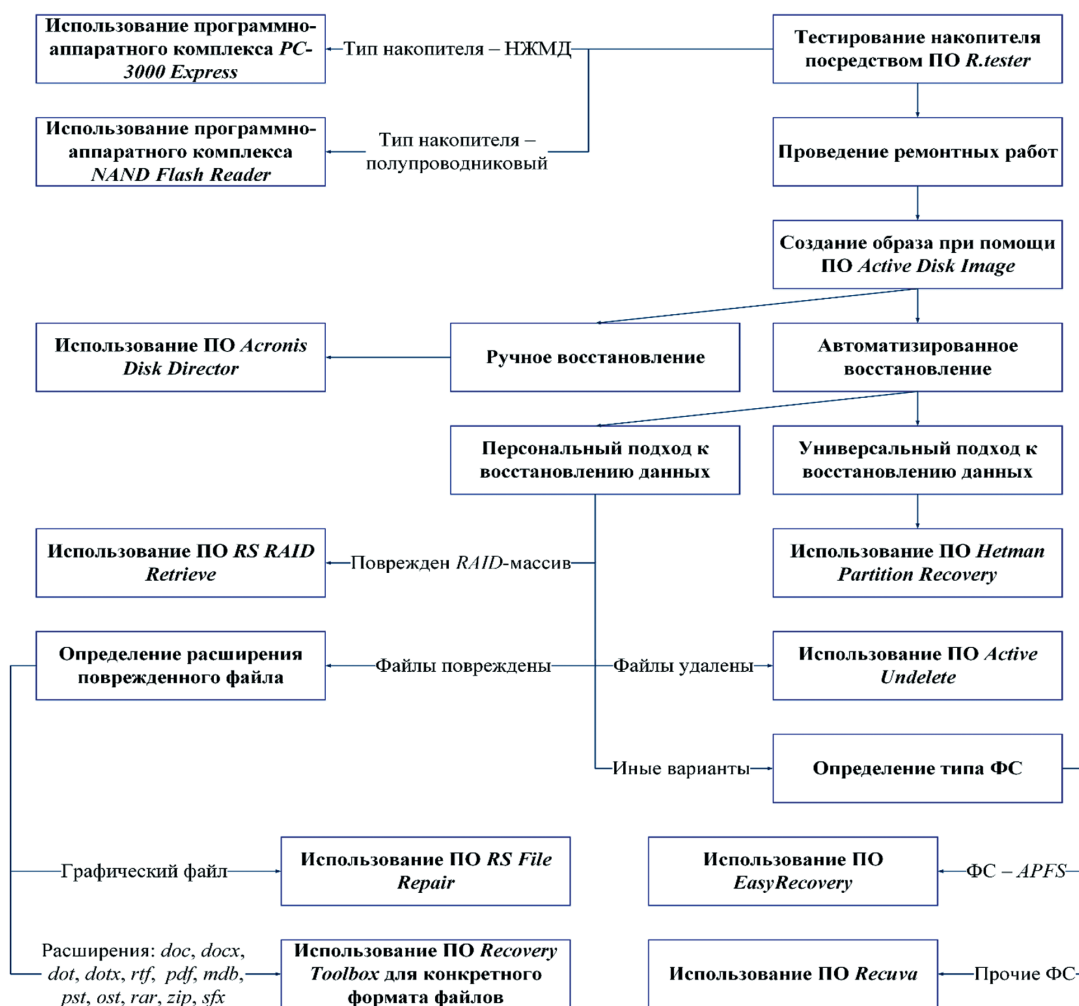


Рис. 4. Методика восстановления информации

$$\text{НОРМ}_x = \text{СУММ}_x \times \text{ВЕКТОР}_x, \quad (5)$$
 где НОРМ_x – нормализованный вектор суждения x ;
 СУММ_x – сумма значений суждений x (по столбцу);
 ВЕКТОР_x – значение вектор-приоритета критерия/ПО x .

$$\text{ИС} = \frac{\text{СУММ}_{\text{НОРМ}} - n}{n - 1}, \quad (6)$$

где ИС – индекс согласованности оценок критериев/оценок ПО;

$\text{СУММ}_{\text{НОРМ}}$ – сумма нормированных векторов суждений;

n – количество критериев или ПО.

$$\text{ОС} = \frac{\text{ИС}}{\text{СЛУЧСОГЛ}} \times 100\%, \quad (7)$$

где ОС – отношение согласованности;

ИС – индекс согласованности;

СЛУЧСОГЛ – мера случайной согласованности.

Рассмотрим суть МАИ на примере выбора ПО для тестирования накопителей. В начале оцениваются критерии, характерные для тестирующих программ

(табл. 2). Далее программы оцениваются по каждому из критериев, например, по результативности (табл. 3).

5. Методика восстановления данных

В результате получилась следующая методика восстановления данных на запоминающих устройствах (рис. 4).

Отдельной группой идут программно-аппаратные комплексы, которые рекомендуется использовать в случае значительных аппаратных проблем накопителей. ПО для ручного восстановления информации необходимо использовать в случае незначительных логических сбоев. Универсальный подход к восстановлению данных реализован с использованием ПО *Hetman Partition Recovery*, а персональный – совокупностью программ в нижней части рис. 4.

6. Тестирование методики

Тестирование разработанной методики проводилось путем решения инцидентов, связанных с утратой до-

Табл. 4. Модели инцидентов

№ инцидента	Накопитель	Файловая система	Причины инцидента	Проявление инцидента
1	НЖМД	NTFS	Удар по НЖМД во время работы	Повреждение БМГ
2	SSD	NTFS	Электрический пробой	Повреждение контроллера
3	Флеш-диск	NTFS	Удаление MBR (замена значений на 0)	Флеш-диск не определяется системой
4	Флеш-диск	NTFS	Удаление файлов на уровне ФС	Файлы недоступны
5	Флеш-диск	NTFS	Значения основных заголовков файлов заменены на 0	Файлы не читаемы
6	НЖМД	NTFS	Несоответствие MFT и битовой карты (значения, соответствующие адресам файлов в MFT заменены на 0)	Файлы недоступны
7	SSD	APFS	Замена значений на 0: загрузочного блока контейнера, заголовков диска, а также удаление файлов на уровне ФС	Диск определяется некорректно

Табл. 5. Экономические модели вариантов применения разработанной методики восстановления данных

№ экономической модели	№ инцидента	Используемые средства	Стоимость технического оснащения, руб.	Стоимость эксплуатации, руб.
1	1	ПО <i>R.test</i> er, Программно-аппаратный комплекс <i>PC-3000 Express</i>	191 900	4 841,58
2	2	ПО <i>R.test</i> er, Программно-аппаратный комплекс <i>NAND Flash Reader</i>	238 243,18	4 024,94
3	3	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>Acronis Disk Director</i>	54 709,82	8 494,19
4	-	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>RS RAID Retrieve</i>	59 318,13	3 298,1
5	4	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>Active Undelete</i>	53 264,11	4 068,11
6	5	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>RS File Repair</i>	52 938,13	2 869,55
7	5	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>Recovery Toolbox</i>	53 770,84	2 869,55
8	-	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>EasyRecovery</i>	57 115,37	2 896,34
9	6	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>Recuva</i>	52 619,13	3 298,1
10	7	ПО <i>R.test</i> er, ПО <i>Active Disc Image</i> , ПО <i>Hetman Partition Recovery</i>	54 818,13	4 101,62

ступа к данным (табл. 4). 1-й и 2-й инциденты предусматривают использование программно-аппаратных комплексов, 3-й инцидент – ПО для ручного восстановления информации. В 4-6-м инцидентах использовалось персональное ПО для восстановления информации, а в 7-м – универсальное.

Все инциденты были успешно решены. Персональное ПО дало самый высокий по скорости результат, что позволяет рекомендовать его для восстановления информации в большинстве случаев.

В табл. 5 приведено соответствие экономических моделей, которые отражают варианты применения методики восстановления данных к инцидентам, на которых тестировалась разработанная методика. В стоимость технического оснащения входят стоимости оборудования и программ. Стоимость эксплуатации зависит от затраченного времени и электроэнергии. Благодаря быстрой работе персонального ПО (экономические модели 4-9) наблюдается экономия денежных средств при эксплуатации.

Заключение

Согласно проведенным работам и исследованиям была создана методика восстановления информации на запоминающих устройствах (см. рис. 4).

Согласно результатам тестирования и расчетам экономических моделей авторами статьи рекомендуется использование персонального ПО для восстановления информации.

Библиографический список

1. Словари и энциклопедии на Академике. Восстановление данных. [Электронный ресурс]. URL: <https://dic.academic.ru/dic.nsf/ruwiki/1369930> (дата обращения: 02.09.2021).
2. Storelab. Что такое восстановление данных? [Электронный ресурс]. URL: <https://storelab-rc.ru/vosstanovlenie-dannih.html> (дата обращения: 02.09.2021).
3. YouTube. Как работает жесткий диск. Что внутри HDD и SSHD. [Электронный ресурс]. URL: <https://www.youtube.com/watch?v=77Ugghti1b0> (дата обращения: 13.09.2021).

youtube.com/watch?v=77Ugghti1b0 (дата обращения: 13.09.2021).

4. Сенкевич Г.Е. Искусство восстановления данных. СПб.: БХВ-Петербург, 2011. 304с.

5. L-pro. Диск формата DVD. [Электронный ресурс]. URL: <http://l-pro.com/uslugi/tirazhirovanie-cd-dvd-diskov/disk-formata-dvd> (дата обращения: 07.09.2021).

6. CNews. Утрата информации ведет к потере бизнеса. [Электронный ресурс]. URL: https://www.cnews.ru/reviews/itinfrastruktura_predpriyatiya_2013/interviews/aleksandr_frolikov (дата обращения: 26.11.2021).

7. Саати Т. Принятие решений. Метод анализа иерархий. М.: Радио и связь, 1989. 316 с.

8. Коробов В.Б., Тутьгин А.Г. Преимущества и недостатки метода анализа иерархий // Известия РГПУ им. А.И. Герцена. 2010. № 122. URL: <https://cyberleninka.ru/article/n/preimuschestva-i-nedostatki-metoda-analiza-ierarhiy> (дата обращения: 20.01.2022).

Сведения об авторах

Василий Сергеевич Замолоцких – обучающийся, Российский университет транспорта, Москва, Российская Федерация, e-mail: netaonh11@gmail.com.

Валентина Геннадьевна Сидоренко – доктор технических наук, профессор, профессор кафедры «Управление и защита информации», Российский университет транспорта, Москва, Российская Федерация, e-mail: valenfalk@mail.ru

Вклад авторов в статью

Замолоцких В.С. Разработка методики восстановления информации на запоминающих устройствах, составление тестирующих моделей инцидентов, тестирование и анализ результатов, формирование выводов.

Сидоренко В.Г. Анализ методов и принципов восстановления информации, обзор методов восстановления информации на запоминающих устройствах.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.