УДК 681.326

А.М. Шашлов

ЭФФЕКТИВНОЕ ВОССТАНОВЛЕНИЕ СВЕДЕНИЙ О РАЗДЕЛАХ ПРИ ПОВРЕЖДЕНИЯХ СИСТЕМ РАЗДЕЛОВ НАКОПИТЕЛЕЙ

Рассмотрены существующие подходы восстановления данных при логических повреждениях систем разделов накопителей. Предложен новый алгоритм восстановления данных, который работает корректно при наличии в адресном пространстве накопителя логических элементов, похожих на логические элементы системы разделов, корректных по формату, но не относящихся к системе разделов накопителя.

E-mail: anthon2k@mail.ru

***Ключевые слова*:** *системы разделов, восстановление данных, логические повреждения.*

Повреждения систем разделов на магнитных и твердотельных накопителях относятся к наиболее частым, но в то же время благоприятным по возможности восстановления информации видам логических повреждений. Такие системы описывают логическую конфигурацию накопителя, состав и параметры логических разделов, на которые он разделен.

Виды систем разделов. Рассмотрим следующие наиболее распространенные виды систем разделов, основанные [1]:

— на таблицах разделов (унаследована от дисковой операционной системы);

— на единой таблице разделов, обозначаемых уникальными идентификаторами (GUID Partition Table), разработана компанией Intel в рамках спецификации Extensible Firmware Interface (EFI), размеры и адреса начала разделов описываются 64-разрядными числами.

Подходы к восстановлению систем разделов. Существующие подходы автоматизированного восстановления систем разделов основаны на сигнатурном поиске в адресном пространстве накопителя сохранившихся логических элементов систем разделов и файловых систем, их верификации и построении на их основе промежуточной таблицы, описывающей разделы, найденные в адресном пространстве накопителя [2].

Последующее восстановление структуры системы разделов проводится, как правило, путем перезаписи в адресном пространстве накопителя логических элементов системы разделов. При этом записываемые логические элементы системы разделов по размещению и содержащимся в них данным могут не полностью соответствовать исходной структуре системы разделов, но обеспечивать возможность доступа к найденным логическим разделам.

Альтернативный подход — предоставление пользователю автоматизированных программных средств восстановления данных возможности копирования информации на другой накопитель.

Необходимо отметить, что адресное пространство накопителя, как правило, содержит фрагменты управляющих логических структур, которые не относятся к системе разделов накопителя (остались от предыдущих логических разбиений, находятся в файлах драйверов и системных утилит или были записаны в файл подкачки или файл спящего режима при работе операционной системы) [3]. Такие логические структуры могут быть корректными по формату, а в ряде случаев соответствовать друг другу и образовывать логически корректные системы разделов (характерно для случаев, когда структуры остаются от ранее существовавших разделов). Проведенное моделирование повреждений показало наличие указанного недостатка во всех рассмотренных в статье программных средствах автоматизированного восстановления данных (в случае, если присутствуют логические структуры, соответствующие по формату таблице разделов и загрузочному сектору одного из разделов). В связи с этим актуальна задача разработки алгоритма восстановления систем разделов, устойчивого к наличию в адресном пространстве накопителя фрагментов управляющих логических структур, которые не относятся к системе разделов накопителя.

Разработка алгоритма восстановления систем разделов. Разработаем алгоритм формирования и оценки гипотез структуры разделов накопителя для восстановления сведений о разделах при повреждении систем разделов накопителя или смешанных повреждениях систем разделов и файловых систем разделов накопителя.

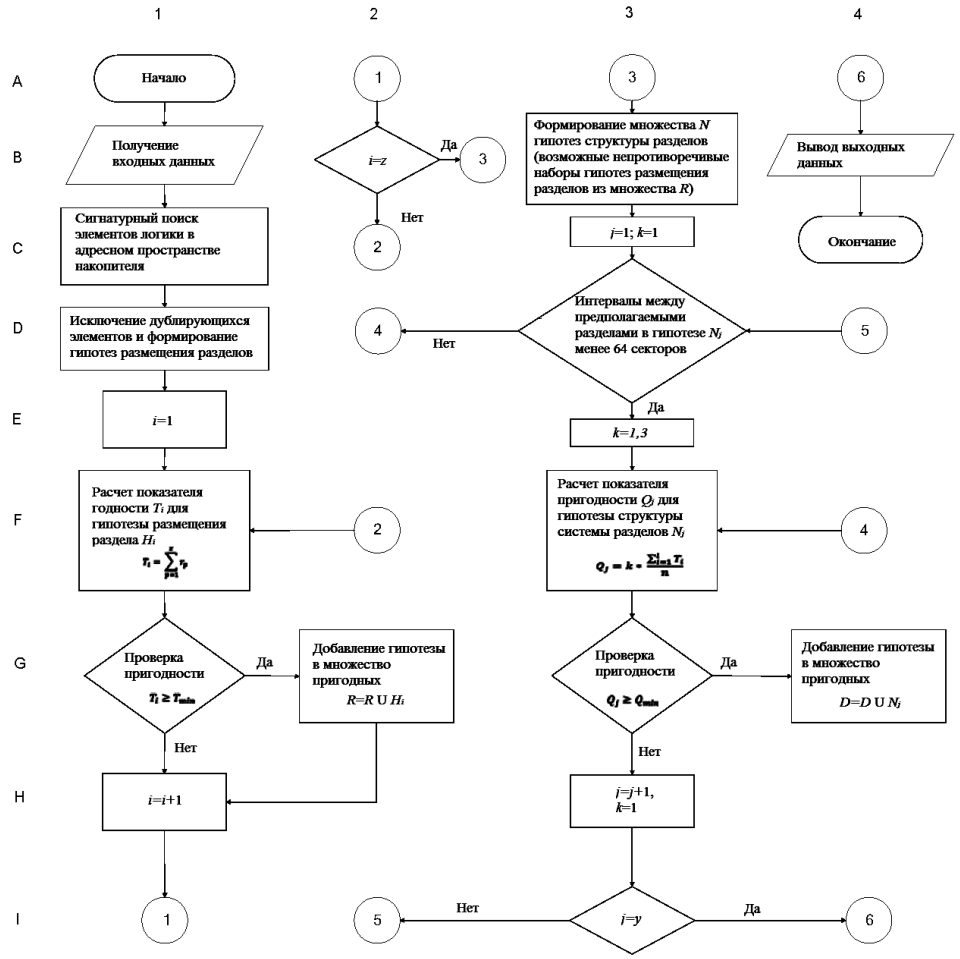
Входные данные для алгоритма: $n_{\text{н}}$ число секторов накопителя; g — размер сектора (число байт в секторе); P — множество секторов адресного пространства накопителя, $P_i = p(i), i = \overline{0, n_{\text{н}}}$ данные (упорядоченные множества байтов), хранимые в каждом из секторов адресного пространства накопителя; $T_{\min}$ — минимально допустимое значение показателя пригодности T_i для гипотез размещения разделов; $Q_{\min}$ — минимально допустимое значение показателя пригодности Q_j для гипотез структуры разделов магнитного и твердотельного накопителей.

Выходные данные для алгоритма: $D_1, \ldots, D_n$ — множество гипотез структуры разделов накопителя, удовлетворяющих требованиям пригодности. При этом каждая гипотеза структуры разделов является непротиворечивым множеством гипотез размещения разделов:

$$D_i = \left(R_1, R_2, \ldots, R_l\right),$$

где l — число гипотез размещения разделов в составе гипотезы структуры системы разделов накопителя, $R_i = r\left(P_i, n_н\right)$, $i = \overline{1, l}$. В ряде случаев носитель может иметь более новые разделы, созданные на месте ранее существовавших, логически корректные, но ценность может представлять именно информация с ранее существовавших разделов.

В связи с указанными факторами предлагаемый алгоритм восстановления структуры разделов предполагает участие эксперта на этапе окончательного выбора гипотезы логического разбиения накопителя на разделы, а выходными данными алгоритма является не одна гипотеза структуры системы разделов, а множество гипотез $D_1, ..., D_n$, удовлетворяющих требованиям пригодности, т. е. таких, на основании которых могут восстанавливаться данные. Рассмотрим блок-схему алгоритма.



Блок-схема алгоритма формирования и оценки гипотез структуры системы разделов

Осуществим поиск логических элементов систем разделов и файловых систем, сохранившихся в адресном пространстве накопителя, путем проверки на наличие известных константных сигнатур во всех секторах носителя [4]. При сигнатурном поиске выявляются следующие, относящиеся к известным элементам системы разделов, либо к известным элементам файловых систем: загрузочный сектор раздела NTFS; загру-

зочный сектор раздела FAT32; первый сектор корневого каталога; сектор GPT; сектор таблицы FAT/FAT32; сектор таблицы MFT.

Кроме сохранившихся элементов и результатов сигнатурного поиска, исходя из значений, которые по умолчанию используют утилиты разбиения диска при формировании блоков зарезервированных секторов между разделами, сектора, находящиеся рядом с обнаруженными разделами, включаются в последующий анализ как возможные места начала и конца логических томов вне зависимости от наличия или отсутствия соответствующих сигнатур.

При анализе структуры разделов найденные ранее базовые элементы (загрузочные сектора — основные и резервные, корневые каталоги, таблицы FAT32 и MFT и иные элементы файловых систем) анализируются на наличие дубликатов. Предполагаемые разделы заносятся в специализированную промежуточную таблицу.

Гипотезы размещения разделов формируются на основе найденных на магнитном или твердотельном накопителе базовых элементов логической структуры системы разделов и файловых систем. Каждая из сформированных гипотез размещения разделов подлежит дальнейшему анализу на предмет возможности восстановления информации.

Для оценки гипотез анализируется каждый из предполагаемых разделов и вычисляется оценочное значение T_i которое определяется суммированием инкрементных значений r_p в соответствии с оценочными шкалами:

Найдены:	
загрузочный сектор (для каждой копии)	2
таблица, определяющая очередность следования кластеров (FAT/MFT):	
для первой копии	14
для второй копии	18
корневой каталог	8
каталог	1
файл	1

Следовательно, оценочное значение T_i для каждой из гипотез размещения раздела в адресном пространстве накопителя рассчитывается по формуле

$$T_i = \sum_{p=1}^{z} r_p,$$

где z — число выявленных элементов логической структуры.

Оценка пригодности каждой из ранее сформированных гипотез R_i размещения раздела проводится сравнением соответствующего данной гипотезе оценочного значения T_i пороговым значением $T_{\min}$. Если $T_i < T_{\min}$, то соответствующая гипотеза R_i исключается из анализа. Таким образом, на этом шаге алгоритма имеем следующий критерий пригодности гипотез R_i размещения разделов:

$$T_i \geq T_{\min}.$$

Гипотезы структуры системы разделов строятся на основе непротиворечивых (отсутствие пересечений между разделами) комбинаций сформированных ранее и прошедших оценку пригодности гипотез размещения разделов. Каждая гипотеза структуры разделов накопителя представляет собой множество гипотез размещения разделов в адресном пространстве накопителя.

Расчет показателя пригодности для гипотезы структуры разделов D_i осуществляется на основе суммы показателей пригодности, определенных ранее для гипотез размещения разделов, входящих в ее состав и специального показателя пригодности (коэффициента поправки k) для гипотез структуры разделов накопителя. Тогда показатель пригодности

$$Q_j = k\frac{\sum_{i=1}^{l} T_i}{n},$$

где k — коэффициент поправки, определяемый с учетом взаимного расположения разделов накопителя; $k = 1$, если расстояния (выраженные в числе секторов) между разделами превышают значения, установленные спецификациями; k = 1, 3, если расстояния между разделами накопителя соответствуют спецификациям.

Окончательный выбор гипотезы разбиения накопителя и принятие решения по восстановлению данных может осуществляться только оператором, восстанавливающим эти данные [5]. В числе прочих причин такое решение обусловлено и тем, что при решении задачи восстановления после некорректных действий персонала носитель может иметь более новые разделы, созданные на месте ранее существовавших, логически корректные, но ценность может представлять именно информация с ранее существовавших разделов. Следовательно, решение об использовании при восстановлении той или иной гипотезы о разбиении диска на разделы оставлено за оператором.

Принимая решение о выборе гипотезы разбиения диска, оператор может ориентироваться на информацию, предоставляемую программой.

Для принятия решения оператору для каждой гипотезы разбиения диска предоставляется следующая информация:

— адреса начала и конца для каждого раздела;

— предполагаемая информация о файловой системе на каждом из разделов;

— краткая оценка состояния раздела.

Если предоставленной информации недостаточно для принятия решения, то возможно проведение нескольких попыток восстановления информации исходя из различных гипотез разбиения диска на разделы. В таком случае оператор сможет принимать решение непосредственно по именам файлов, которые подлежат восстановлению.

При невозможности принятия решения по выбору той или иной гипотезы логического разбиения диска оператор может проанализировать структуру каталогов для одного или нескольких предлагаемых вариантов. Для анализа структуры каталогов потребуется чтение области данных предполагаемых разделов и время, зависящее от размера раздела.

После уточнения предполагаемой гипотезы разбиения диска на логические разделы для каждого раздела проводится полный анализ сохранившихся элементов файловой системы, создается промежуточная таблица, определяющая размещение файлов на диске и последовательность кластеров, составляющих каждый из обнаруженных файлов. На этом шаге также выявляются кластеры, которые не входят в файлы.

Заключение. Предложен алгоритм формирования и оценки гипотез структуры разделов в адресном пространстве накопителя. Алгоритм отличается от аналогов устойчивостью к наличию в адресном пространстве накопителя фрагментов управляющих логических структур, которые корректны по формату, но не относятся к исходному разбиению диска на разделы (остались от предыдущих логических разбиений, находятся в файлах драйверов и системных утилит, либо попали в файл подкачки или файл спящего режима при работе системы).

СПИСОК ЛИТЕРАТУРЫ

1. Russel Ch., Craftword Sh., Gerend J. Microsoft Windows Server 2003 Administrator's companion. — Redmond, WA: Microsoft Press, 2003. — 1632 p.
2. Tyagi T. Data Recovery with and without Programming // BPB Publications, November 2004. — 540 p.
3. Ховард М., Лебланк Д. Защищенный код. Пер. с англ. — М.: Издательско-торговый дом «Русская Редакция», 2004. — 704 с.
4. Custer H. Inside the Windows NT File System. — Redmond, WA: Microsoft Press, 1994. — 91 p.
5. Обеспечение безопасности информации в центрах обеспечения управления полетами космических аппаратов / Л.М. Ухлинов, М.П. Сычев, В.Ю. Скиба и др. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2000. — 366 с.

Статья поступила в редакцию 4.07.2012