СНИЖЕНИЕ ОБЪЕМА ОБРАБАТЫВАЕМОЙ ИНФОРМАЦИИ В ЭНЕРГОЗАВИСИМОЙ ПАМЯТИ ПРИ ИССЛЕДОВАНИИ КОМПЬЮТЕРНЫХ ИНЦИДЕНТОВ

Пантюхин И.С.[1], Белов Н.И.[2], Катаева В.А.[3]

Предложен метод снижения объема обрабатываемой информации для расследования киберинцидентов в энергозависимой памяти. Сущность предлагаемого решения заключается в определении и исключении мало информативных данных, содержащихся в дампах оперативной памяти. Метод обеспечивает снижение объема данных с минимальными потерями информации, являющихся важными при расследовании киберинцидентов. Первый этап работы метода – подготовка входных данных для анализа, он заключается в получении и формировании дампов оперативной памяти компьютера. Далее происходит разбор этих данных на набор атрибутов и последующая классификация некоторых значений атрибутов, составляющих основу для проведения постинцидентного анализа в энергозависимой памяти. Полученные данные структурируются и записываются в CSV файл. На втором этапе происходит расчет информативности каждого из атрибутов с помощью метода Шеннона. На заключительном этапе происходит исключение атрибутов, классифицированных как малоинформативные. Предложенный метод позволяет существенно снизить временные затраты и объем данных, занимаемыми дампами памяти, не потеряв при этом важную информацию для проведения внутреннего аудита в энергозависимой памяти.

***Ключевые слова**: метод, снижение объема, информативность, метод Шеннона, энергозависимая память, компьютерная криминалистика*

DOI: 10.21681/2311-3456-2018-2-70-76

Введение

С каждым годом растет объем данных, обрабатываемых в информационных системах, вместе с ним стремительно растет количество инцидентов информационной безопасности, в частности – в энергозависимой памяти. В связи с этим появляются определенные задачи в области расследования инцидентов, такие как изучение существующих способов получения дампов энергозависимой памяти, исследование существующих методов расчета информативности признаков и разработка метода, способного уменьшить объем, занимаемый дампами энергозависимой памяти на диске, путем исключения данных, которые не являются необходимыми или полезными для расследовании киберинцидентов. Таким образом, осуществляя исключение данных, которые не несут в себе важной информации, можно снизить количество памяти, занимаемое дампом, которое необходимо для проведения внутреннего аудита большого объема данных оперативного запоминающего устройства (далее – ОЗУ).

На сегодняшний день авторам не известен метод, применяемый в сфере информационной безопасности и форензики, который позволил бы существенно уменьшить объем информации, занимаемый дампами энергозависимой памяти, без весомых потерь информативности этих данных.

Методы и материалы

Метод снижения объема обрабатываемой информации для расследования киберинцидентов в энергозависимой памяти включает несколько последовательных шагов. На первом шаге осуществляется процесс получения и формирования данных, которые состоят из дампа оперативной памяти (RAM) [1]. Для полученных данных была определена следующая структура. После этапа предобработки атрибуту, имеющему в себе значение пути к процессу, было присвоено одно из значений класса, основанное на уровне вложенности файла в каталогах [2]. Для хранения и удобства обработки полученные структурированные данные были записаны в CSV файл. Алгоритм работы данного метода представлен на рис. 1.

Для получения данных с дампов ОЗУ использовалась утилита volatility framework, написанная на языке Python [3]. В результате этапа была получена таблица, где каждая строка – определенный процесс, а столбец – атрибут данного процесса, например: название процесса (Name), идентификатор процесса (UID), идентификатор родительского процесса (PPID), дата и время начала работы

1 Пантюхин Игорь Сергеевич, ассистент, Университет ИТМО, Санкт-Петербург, Россия.E-mail: zevall@cit.ifmo.ru
2 Белов Никита Игоревич, магистрант, Университет ИТМО, Санкт-Петербург, Россия. E-mail: nikit.belov@gmail.com
3 Катаева Валентина Алексеевна, магистрант, Университет ИТМО, Санкт-Петербург, Россия.E-mail: kataeva@cit.ifmo.ru

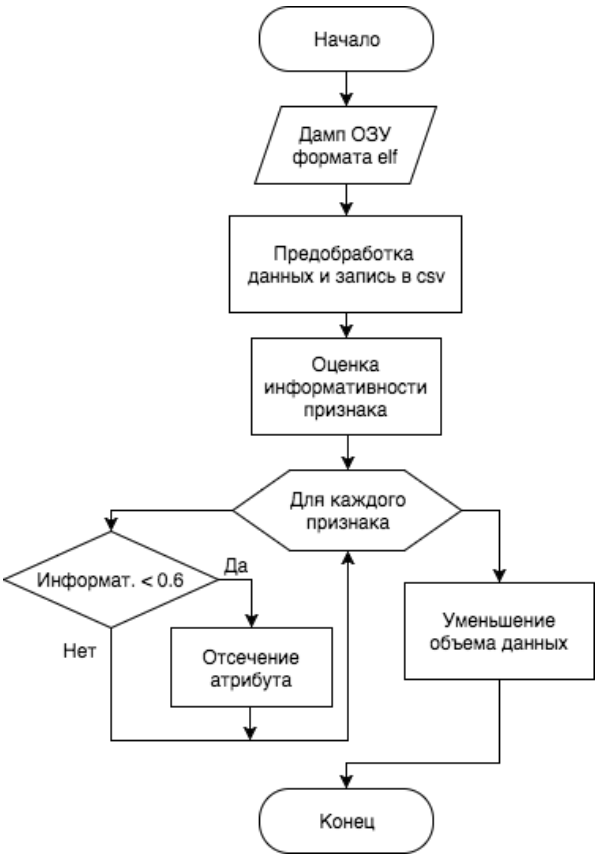

***Рис. 1**. Блок-схема работы предлагаемого метода*

процесса (Start-date-time), использование 64 битное или 32 битное пространство адресов (Wow64) и др. [4]. Эти атрибуты и их значения для каждого процесса сохраняются в CSV файле и имеют структуру, показанную на рис. 2.

Также в полученной таблице для дальнейшего расчета информативности атрибута был создан столбец, в котором записаны значения 0 и 1, обозначающие принадлежность данного процесса к инциденту. Для уменьшения вероятности ошибки подпроцессы, вызванные процессами, относящимися к инцидентам, также считались причастными к возникновению инцидентов информационной безопасности. Для наглядности было построено дерево процессов, где красным выделены вредоносные процессы, оно представлено на рис. 3.

Далее для определения того, какие атрибуты являются информативными в целях идентификации процесса, вызвавшего возникновение киберинцидента в оперативной памяти, необходимо рассчитать информативность каждого атрибута. Это возможно сделать с помощью различных методов, например, метода Шеннона, метода Кульбака или метода Накопленных Частот (НЧ). Сравнительный анализ данных методов представлен в табл. 1 [5]. Было принято решение использовать

Offset (P)	Offset (V)	Name	PID	PPID	PDB	Start-time	Exit-time	Wow64	Path
0x000000010f7da854	0xfffffa80045c2060	cmd.exe	4708	1564	0x0000000077a88000	11:15:55	0	1	3
0x000000011039b324	0xfffffa8003982b30	audiodg.exe	4352	772	0x000000006d381000	11:06:36	0	0	3
0x000000010e59c324	0xfffffa8005783b30	SearchIndexer	1328	484	0x0000000079465000	8:01:44	10:09:32	0	3
0x00000001102ea854	0xfffffa80038d2060	msiexec.exe	3224	484	0x0000000107786000	10:06:28	0	0	3
0x000000010ff80bb4	0xfffffa8003d683c0	opera.exe	3288	2156	0x0000000010d5e000	8:25:28	0	1	0
0x000000010e5f21d4	0xfffffa80057d99e0	svchost.exe	1940	484	0x000000007dc39000	8:01:30	0	0	3
0x000000010fe4b0c4	0xfffffa8003c328d0	cmd.exe	4956	1564	0x0000000050c23000	11:07:16	0	1	3
0x000000011023de94	0xfffffa80038256a0	svchost.exe	2472	484	0x0000000048cc6000	10:05:09	0	0	3
0x0000000007148f74	0xfffffa8003744780	file.exe	5032	3240	0x000000002faee000	11:16:17	0	1	7
0x000000010ffb2324	0xfffffa8003d99b30	opera.exe	3984	2156	0x00000000ac883000	9:20:30	0	1	0

***Рис. 2**. Пример структуры хранения атрибутов данных и их значений в CSV файле*

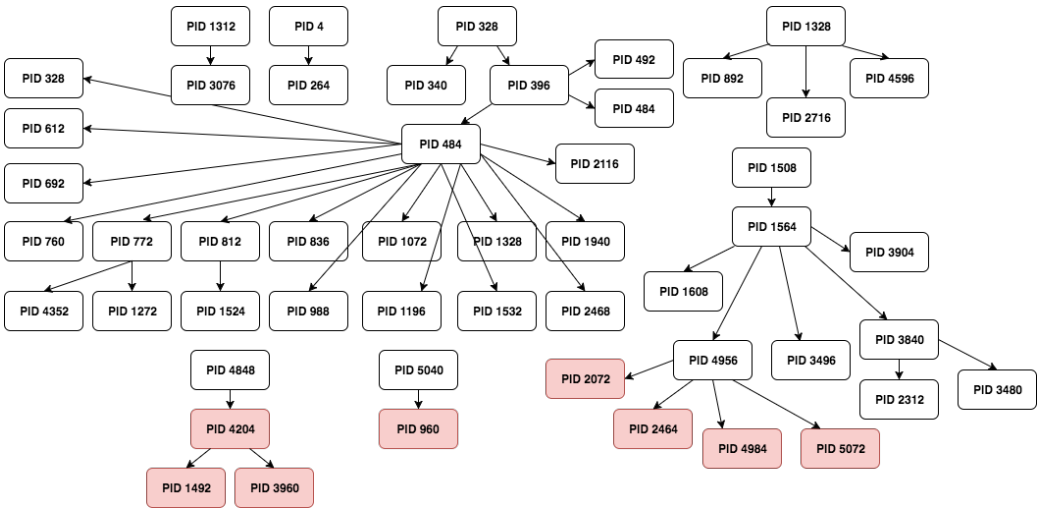

***Рис. 3**. Дерево процессов одного из дампов ОЗУ*

Таблица 1
Сравнительный анализ методов оценки информативности

Критерии	Метод Шеннона	Метод НЧ	Метод Кульбака
Зависимость метода от способа кодировки атрибута	Не зависит	Зависит	Не зависит
Число классов	Произвольное число классов	Два класса	Два класса
Зависимость метода от объема выборки	Не зависит	Зависит	Не зависит
Нормированность результата	Нормирована	Не нормирована	Не нормирована

Метод Шеннона для данной задачи по следующим причинам:

• Данный метод не зависит от способа представления значения признака, в нашем случае данные могут быть как текстового типа, так и числового.

• Результатом расчета информативности с помощью метода Шеннона является нормированная величина от 0 до 1. Поэтому значение информативности близкое к 1 можно считать высоким, а значение близкое к 0, соответственно, низким.

На заключительном этапе проводится классификация атрибутов по значению их информативности. Для проведения классификации атрибутов необходимо выбрать пороговое значение информативности, при использовании которого будут определены необходимые для расследования атрибуты и исключены малоинформативные, а также обеспечена достаточная точность классификации атрибутов с целью их использования для расследования киберинцидентов. Был построен график, показывающий вероятность возникновения ошибки, т.е. исключения важного для постинцидентоного анализа атрибута, используя шкалу вероятности от 0 до 100% и шкалу значений информативности атрибутов от 0 до 1. Опытным путем было выбрано пороговое значение информативности – 0,6, т.к. данное значение не превышает выбранный нами максимально-приемлемый критерий ошибки равный 5%, и в тоже время оно дает максимальное снижение объема данных. Данный график представлен на рис. 4.

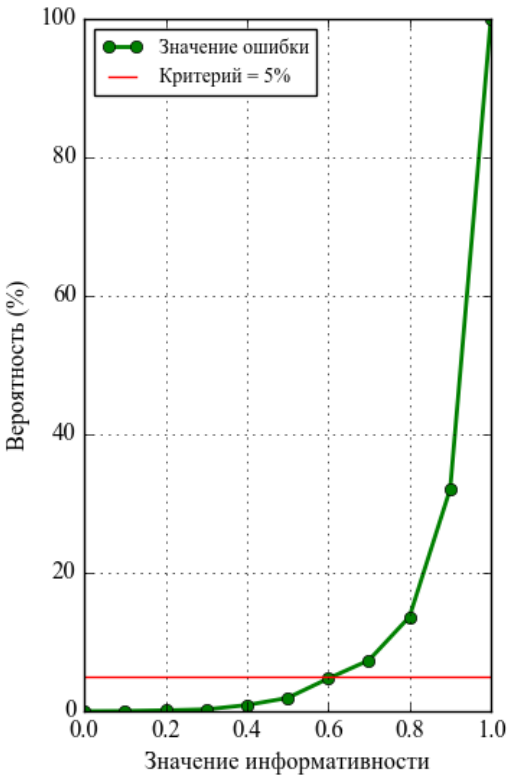

Рис. 4. *Вероятность возникновения ошибки классификации*

На последнем этапе была повторно проведена классификация атрибутов. Атрибуты, имеющие информативность большую или равную выбранному пороговому значению, были переклассифицированы как "Информативные", а те атрибуты, которые имели значение информативности меньше порогового значения, – как "Не информативные", и были исключены.

Метод Шеннона

Один из способов оценки информативности известный из теории информации – метод Шеннона – использует для оценки информативности средневзвешенное количество информации, которое приходит на разные градации признака. Информацией в теории информации принято понимать величину устраненной энтропии. Метод Шеннона основан на вероятностях, поэтому объемы выборки наблюдений признака по трем распознаваемым классам может быть различен, в

Таблица 2
Значения информативности атрибутов

Название атрибута	Значение информативности
Offset(P)	0.36238227890860503
Offset(V)	0.79696350463327903
Name	0.45162845283631903
UID	0.51296684556078098
PPID	0.52295301466869104
PDB	0.64644890872275695

отличии от Метода накопленных частот, так как он оперирует частотами, и объем выборки наблюдений признака должен быть одинаков по распознаваемым классам [6].

На тестовой выборке следующим образом были определены процессы, относящиеся к инцидентам: на основе сравнения двух дампов (до появления инцидента в системе и после) процессы, которых не было в первом дампе, считались относящимися к инцидентам, и им был присвоен ранг 1, остальным процессам – 0. После получения и формирования дампов оперативной памяти, как говорилось ранее, происходил разбор этих данных на атрибуты, и поочередный расчет значения информативности каждого атрибута для последующей классификации. В методе Шеннона рассчитывается информативность для каждого j-го признака, под признаками в данном случае понимаются атрибуты [2].

Информативность j-ого признака рассчитывается по формуле (1).

$$I(x_i) = 1 + \sum_{i=1}^{G}(P_i \cdot \sum_{k=1}^{K} P_{i,k} \cdot log_k P_{i,k}) \quad (1)$$

Где G – количество градаций признака, K – количество классов, P_i – вероятность i-той градации признака, которая считается по формуле (2).

$$P_i = {\sum_{k=1}^{K} m_{i,k}}/{N} \quad (2)$$

Где $m_{i,k}$ – частота появления i-той градации в K-том классе, N – общее число наблюдений. $P_{i,k}$ – вероятность появления i-той градации признака в K-том классе, она считается по формуле (3).

$$P_{i,k} = \frac{m_{i,k}}{\sum_{k=1}^{K} m_{i,k}} \quad (3)$$

Метод Шеннона оценивает информативность, как нормированную величину, которая изменяется в промежутке от 0 до 1. Таким образом, информативность признака, определяемая данным методом рассматривается в абсолютном плане: ближе к 1 – высокая; ближе к 0 – низкая [7].

Таким образом была рассчитана информативность для каждого атрибута. Результаты расчетов информативности методом Шеннона для некоторых атрибутов представлены в табл. 2.

Результаты

В качестве объектов исследования был создан экспериментальный стенд, состоящий из компьютеров (10 шт.) следующей конфигурации:

Процессор	Intel Core i7 2600K 3.4Ghz
Оперативная память	3 ГБ
Операционная система	Windows 10 x64

Каждый компьютер выступал в качестве средства вычислительной техники для проведения постинцидентного анализа экспертом. Была разработана программная реализация предложенного метода снижения объема обрабатываемой памяти для расследования киберинцидентов в энергозависимой памяти. Основой метода является определение и исключение атрибутов процессов памяти, которые являются неинформативными при расследовании компьютерных инцидентов [7].

Всего было проведено 8 экспериментов на разных дампах ОЗУ. Результаты экспериментов, полученных с помощью программной реализации данного метода, представлены на рис. 5 - рис. 6. На рис. 5 представлено количество малоинформативных атрибутов, отнесенных к классу “Не информативные”, и количество важных для расследования атрибутов, отнесенных к классу “Информативные”, определенных в результате проведения одного из экспериментов.

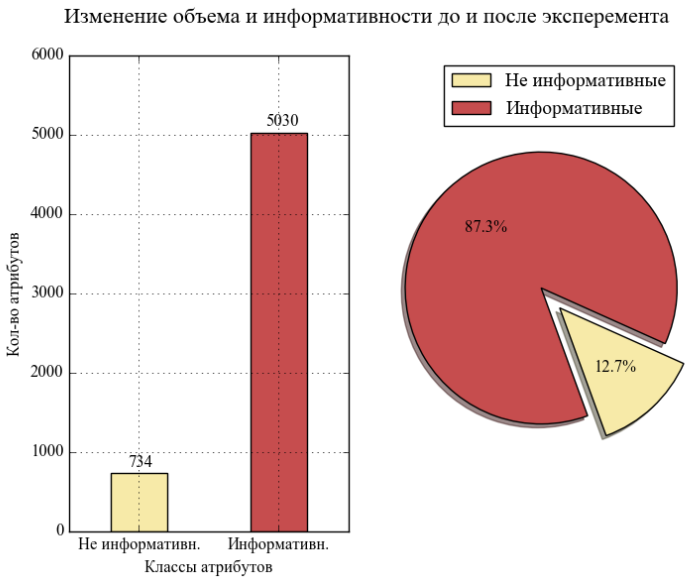

Рис. 5. *Количество атрибутов в каждом классе*

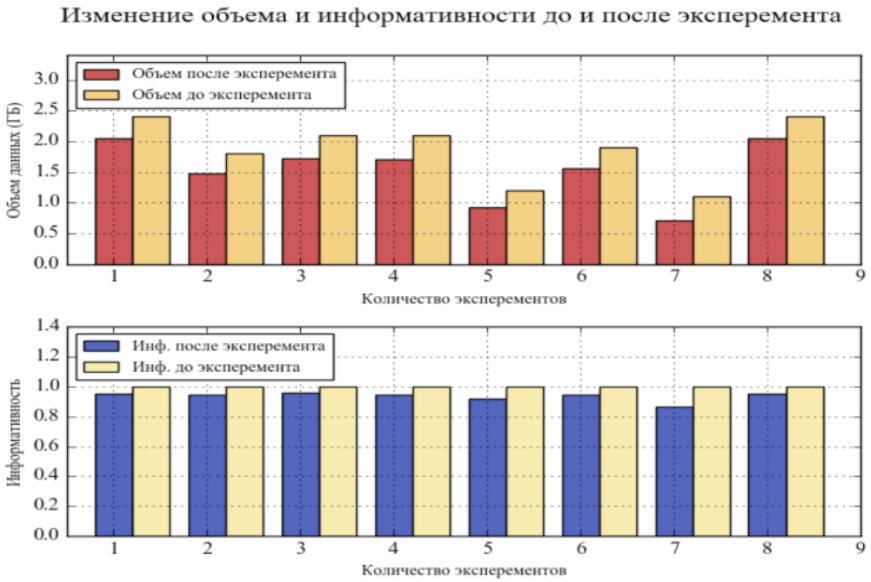

Рис. 6. *Результаты эксперимента*

На рис. 6 показаны объемы обрабатываемых данных до и после применения описанного метода. В среднем, в процессе проведения всех экспериментов объем обрабатываемых данных без потери информативности уменьшился на 22%, что подтверждает эффективность работы данного метода [8].

Заключение

Предложенный метод снижения объема обрабатываемой информации для расследования киберинцидентов в энергозависимой памяти позволяет существенно уменьшить объем, занимаемый данными, необходимыми для проведения внутреннего аудита. Также достоинством представленного метода является то, что он позволяет снизить временные затраты на проведение внутреннего аудита энергозависимой памяти, при этом без потери важной информации для расследования компьютерных инцидентов. Данный метод может быть адаптирован под другие виды операционных систем и компьютерных инцидентов, а также широко использоваться в различных задачах компьютерной криминалистики в совокупности с другими методами.

Литература

1. Limon G.G. Forensic physical memory analysis: an overview of tools and techniques / In: TKK T-110.5290 Seminar on Network Security. Helsinki, Finland, 2007. P. 305–320.
2. Shannon, C.E. The Mathematical Theory of Communication / C.E.Shannon and W.Weaver. – Urbana, IL: University of Illinois Press. – ISBN: 0252725484. – 1963. – 144 p.
3. Analyzing a RAM Image with Volatility [Электронный ресурс] –URL: https://samsclass.info/121/proj/p4-Volatility.htm, режим доступа: свободный, дата обращения 22.03.2017.
4. Kristine Amari, Techniques and Tools for Recovering and Analyzing Data from Volatile Memory. 26 March 2009. 59 P. URL: https://www.sans.org/reading-room/whitepapers/forensics/techniques-tools-recovering-analyzing-data-volatile-memory-33049 (дата обращения: 10.05.2018).
5. Пантюхин И.С., Зикратов И.А., Левина А.Б. Метод проведения постинцидентного внутреннего аудита средств вычислительной техники на основе графов // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 3. С. 506-512.
6. Дашутина Е.В., Меркулова Е.В. Проектирование СКС диагностики сердечно-сосудистых заболеваний // Материалы II международной научно-технической конференции студентов, аспирантов и молодых ученых. Донецк, ДонНТУ. 2012. С. 285-288.
7. Kullback S. Information Theory and Statistics. — John Wiley & Sons, 1959.
8. Volatility Foundation – ForensicsWiki website, [Электронный ресурс] / URL: http://www.volatilityfoundation.org/about, режим доступа: свободный, дата обращения 14.03.2017.

REDUCING THE AMOUNT OF PROCESSED INFORMATION IN VOLATILE MEMORY IN INVESTIGATING COMPUTER INCIDENTS

I. Pantiukhin [4], N. Belov [5], V. Kataeva [6]

Abstract. A algorithm for reducing the amount of processed information for investigating cyber incidents in volatile memory is proposed. The key points of the proposed solution are determining and excluding non–informative data in random access memory (RAM) dumps. The method ensures a reduction of the amount of data with minimal data losses important for investigating cyber incidents. At the first stage of the method's work, input data for an analysis is prepared which consists in getting and forming RAM dumps. After that, an analysis of the data in order to identify an attribute array is carried out with a subsequent categorisation of some values of the attributes which form the basis for carrying out a post–incident analysis in volatile memory. The data obtained are given a structure and written to a CSV file. At the second stage, computations of informativeness of each attribute using Shannon's method are performed. At the final stage, attributes categorised as non–informative are excluded. The proposed method allows to substantially reduce the time expenditures and amount of data taken by the RAM dumps without losing information important for carrying out an internal audit in volatile memory.

***Keywords**: algorithm, amount reduction, informativeness, Shannon's method, volatile memory, computer forensic science.*

4 Igor' Pantiukhin, Assistant Professor, University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia. E-mail: zevall@cit.ifmo.ru

5 Nikita Belov, Master's Student, University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia. E-mail: nikit.belov@gmail.com

6 Valentina Kataeva, Master's Student, University of Information Technologies, Mechanics and Optics, Saint Petersburg, Russia. E-mail: kataeva@cit.ifmo.ru

References

1. Limon G.G. Forensic physical memory analysis: an overview of tools and techniques / In: TKK T-110.5290 Seminar on Network Security. Helsinki, Finland, 2007. P. 305–320.
2. Shannon, C.E. The Mathematical Theory of Communication / C.E.Shannon and W.Weaver. – Urbana, IL: University of Illinois Press. – ISBN: 0252725484. – 1963. – 144 p.
3. Analyzing a RAM Image with Volatility [Electronic resource] –URL: https://samsclass.info/121/proj/p4-Volatility.htm, date of circulation: 22.03.2017.
4. Kristine Amari, Techniques and Tools for Recovering and Analyzing Data from Volatile Memory. 26 March 2009. 59 P. URL: https://www.sans.org/reading-room/whitepapers/forensics/techniques-tools-recovering-analyzing-data-volatile-memory-33049 (date of circulation: 10.05.2018).
5. Pantiuhin I.S., Zikratov I.A., Levina A.B. Metod provedeniia postintcidentnogo vnutrennego audita sredstv vy`chislitel`noi` tekhniki na osnove grafov // Nauchno-tekhnicheskii` vestneyk informatcionny`kh tekhnologii`, mehaniki i optiki. 2016. T. 16. № 3. S. 506-512.
6. Dashutina E.V., Merkulova E.V. Proektirovanie SKS diagnostiki serdechno-sosudisty`kh zabolevanii` // Materialy` II mezhdunarodnoi` nauchno-tekhnicheskoi` konferentcii studentov, aspirantov i molody`kh ucheny`kh. Donetck, DonNTU. 2012. C. 285-288.
7. Kullback S. Information Theory and Statistics. — John Wiley & Sons, 1959.
8. Volatility Foundation – ForensicsWiki website, [Electronic resource] / URL: http://www.volatilityfoundation.org/about, date of circulation: 14.03.2017.