УДК 004.056.52 DOI: 10.18698/2541-8009-2018-8-359

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ПОИСКА СИГНАТУР КОМПЬЮТЕРНЫХ АТАК В СЛОВАРЕ ПРИЗНАКОВ

Е.М. Череданова pankooova@mail.ru
SPIN-код: 1619-5499

И.Р. Печкурова risha.irisha18@mail.ru
SPIN-код: 2108-3379

В.Р. Печкурова vika.florida@gmail.com
SPIN-код: 4792-8460

Е.А. Мамченко liza.98.98@mail.ru
SPIN-код: 2887-3715

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Возможность выявления компьютерных DDoS-атак в реальном масштабе времени является одним из ключевых аспектов защиты информации, обеспечения безопасности сетевых ресурсов, а также ограничения доступа к серверным вычислительным мощностям. В работе проведен анализ эффективности трех видов (последовательный, итерационный и с использованием хеш-адресации) сигнатурного метода, основанного на поиске признаков компьютерной атаки в соответствующем словаре. В результате исследования выявлены достоинства и недостатки всех рассматриваемых методов, а также факторы, влияющие на принятие решения о выборе того или иного вида сигнатурного метода. Это позволило сформировать в обобщенном виде краткие рекомендации специалистам по защите информации и обеспечения безопасности сетевых ресурсов по использованию трех рассмотренных разновидностей метода сигнатур для обнаружения DDoS-атаки.

Ключевые слова

Метод сигнатур, DDoS-атака, словарь признаков, итерация, хеш-функция, последовательный метод, итерационный метод, метод хеш-адресации

Поступила в редакцию 28.05.2018

Методы поиска признаков компьютерных атак. В настоящее время одну из существенных угроз функционированию вычислительных сетей представляют компьютерные атаки (КА). Одной из разновидностей КА является DDoS-атака — атака на вычислительную систему с целью довести ее до отказа, т. е. создание таких условий, при которых легитимные (правомерные) пользователи системы не могут получить доступ к предоставляемым системой ресурсам (серверам), либо этот доступ крайне затруднен. Важными факторами обеспечения эффективного противодействия КА являются их оперативное обнаружение, идентификация и реагирование на подозрительную деятельность по осуществлению доступа к вычислительным или сетевым ресурсам.

Один из основных методов обнаружения КА — использование сигнатур атак. Метод заключается в описании атаки в виде сигнатуры (набора признаков) и поиска данной сигнатуры в контролируемом пространстве. В качестве сигнатуры атаки может выступать шаблон действий или ряд символов, характеризующих аномальную деятельность. Данные сигнатуры представляются в виде идентификаторов (признаков) КА и хранятся в словаре признаков. При использовании метода сигнатур атак процесс поиска идентификатора КА в словаре признаков требует значительных временных ресурсов. Это делает актуальным решение задачи минимизации времени поиска признаков КА благодаря использованию эффективных методов их обнаружения в словаре [1].

Последовательный метод. Одним из методов поиска признаков КА в словаре является последовательный метод. Его главное достоинство заключается в том, что сравнение параметра поиска с признаками КА в словаре осуществляется посимвольно по всей длине (при искажении отдельных символов позволяет принимать то или иное решение по результатам поиска).

Наряду с указанным достоинством последовательный метод обладает и существенным недостатком, связанным со значительными затратами времени на поиск, что при больших объемах словаря признаков не позволяет осуществлять анализ и выявление КА в реальном масштабе времени [2].

Оценим временные затраты (количество итераций) для данного метода поиска. Предположим, что длина параметра поиска составляет M символов, а число признаков в словаре — N. Тогда предельное количество операций (шагов), в результате которых будет проанализирован весь объем словаря, составит

$$R_{\text{посл}} = N \times M.$$

Итерационный метод. Суть итерационного метода заключается в следующем. На первом шаге сравнивается первый символ параметра поиска с первыми символами всех N признаков словаря. Признаки, у которых первый символ не совпал, на втором шаге не анализируются. Иными словами, по второму символу сравниваются лишь те признаки, у которых совпал первый. На третьем шаге анализируются только те признаки словаря, у которых совпали первый и второй символы параметра поиска. На четвертом — первый, второй и третий и т. д., пока не будут проанализированы все символы параметра поиска.

Данная процедура приводит к сокращению времени поиска в словаре и, как следствие, к повышению оперативности обнаружения признаков КА. Скорость сходимости данного метода

$$R_{\text{итер}} = \log_2 N.$$

Это означает, что не более чем через $R_{\text{итер}}$ равнений мы либо найдем признак в словаре, либо убедимся в его отсутствии [3].

Метод хеш-адресации. Существенно сократить время поиска в словаре можно путем разработки специальных мер увеличения скорости поиска, например, благодаря внедрению хеш-функции.

Сущность метода, основанного на хеш-адресации, заключается в том, что на начальном этапе все признаки КА преобразуются в числа, которые соответствуют их адресам в словаре. При последующем анализе комбинация символов параметра поиска через использование хеш-функции преобразуется в число, которое используется в качестве адреса словаря, для выбора и сравнения признака КА (рис. 1).

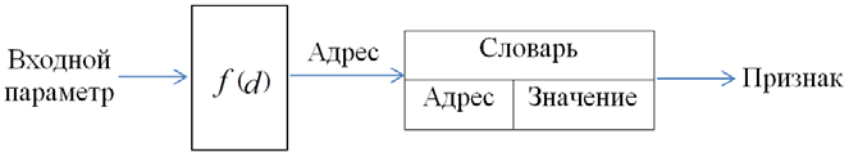

Рис. 1. Метод хеш-адресации

В общем случае под хеш-функцией будем понимать функцию, определенную на множестве комбинаций символов признака и присваивающую ему число, которое является адресом признака в словаре (число строк в словаре полагаем заданным и равным N). Если d означает произвольную структуру комбинации символов признака, i — номер строки словаря, а f — хеш-функцию, то отображение, реализуемое хеш-функцией, условно можно записать следующим образом:

$$i = f(d).$$

Пример хеш-адресов для значений словаря приведен в табл. 1.

Таблице 1

Хеш-адресация значений словаря признаков КА

Хеш-адрес	Значения словаря	
33	cost	range
72	coalition	css
108	asw	afload
115	trigger	—

Предлагаемый метод не требует дополнительных временных затрат на просмотр всего словаря признаков, объем которого может быть практически неограниченным. Процедура хеширования выполняется в реальном масштабе времени, что и является достаточным условием для применения данного метода для решения задач обнаружения признаков КА [4].

Программная реализация методов поиска признаков компьютерных атак в словаре признаков. В качестве условного словаря признаков для моделирования использовался словарь терминов объемом в 395 слов. Предварительно была осуществлена сортировка слов словаря по алфавиту. На основе данных словаря была разработана хеш-таблица словарных значений. Интерфейс разработанного программного обеспечения имитационного моделирования представлен на рис. 2.

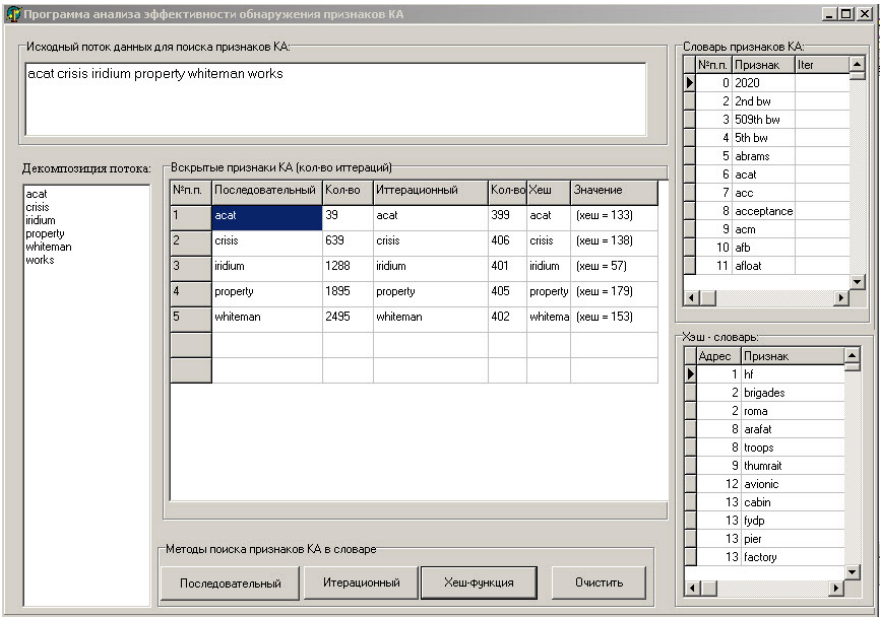

Рис. 2. Интерфейс программы имитационного моделирования

Последовательный поиск реализован процедурами перебора и поиска полного совпадения искомого слова с элементами словаря, начиная с его первого элемента. В качестве дополнительного недостатка при использовании данного метода можно полагать тот факт, что скорость поиска определенного термина в словаре признаков КА в реальных условиях (при его априорно известном наличии), т. е. при возможном отсутствии сортировки элементов словаря по алфавиту, будет зависеть от такого слабо прогнозируемого и стохастического фактора, как положение искомого слова в словаре ближе к его началу или концу [5].

Итерационный поиск, реализующий первый проход сравнения первых символов элементов словаря признаков, аналогичен последовательному алгоритму, однако после первого прохода в дальнейшем поиске участвуют только те элементы словаря, у которых были отмечены совпадения. Аналогичный отбор происходит после второго и последующих проходов. Таким образом, количество проходов (а значит, и временных затрат) существенно сокращается в сравнении с последовательным поиском.

Наиболее быстрым является поиск в словаре через вычисление хеш-функции входного параметра. После вычисления функции осуществляется переход на строку в словаре с номером вычисленной хеш-функции. Если по этому адресу присутствует больше одного признака КА (наблюдаются коллизии), то среди них осуществляется поиск итерационным методом [5].

Реализуем словарь на основе использования хеш-адресации с помощью хеш-таблицы. Для этого хеш-таблицу представим в виде массива, задаваемой хеш-функцией. Размер таблицы признаков КА должен быть достаточно большим, чтобы в ней оставалось достаточно большое число пустых мест. В процессе формирования хеш-словаря признаков для вставки в него нового значения

(признака КА), мы должны предварительно хешировать данное значение (получить число), чтобы определить ключ, который будет соответствовать размещению нового значения в хеш-словаре признаков КА. Для удаления признака из словаря, мы находим его и удаляем элемент словаря по данному хеш-адресу [6].

Для построения хеш-словаря признаков КА и поиска параметра в нем мы должны использовать хеш-функцию. Предположим, что число символов, применяемых при написании структур комбинации признаков, равно m. Несмотря на то что в реальных условиях на вход могут подаваться различных слова на языках, а также цифры и специальные символы, для упрощения примем m равным 26, количеству букв в алфавите английского языка. Через $[d]$ обозначим число, образуемое цифрами, отражающими в некотором порядке символы алфавита в m-ичной системе счисления [7]. В общем виде $[d]$ будет представлять собой полином

$$[d] = s_0 m^{i-1} + s_1 m^{i-2} + s_2 m^{i-3} + \ldots + s_{i-2} m^1 + s_{i-1} m^0,$$

где s_i — кодовое обозначение символа в формате американского стандарта для обмена информацией ASCII (American Standard Code for Information Interchange); i — количество символов (длина) слова.

Тогда в качестве хеш-функции можно, например, взять выражение

$$f(d) = ([d] \bmod N) + 1,$$

где N — число строк словаря.

Рассмотрим детально нахождение хеш-функции на следующем примере.

Исходный параметр поиска (слово) — «acat». Каждый символ представим в коде формата ASCII: символ «a» = 97, символ «c» = 99, символ «a» = 97, символ «t» = 116.

Число символов, применяемых при написании структур комбинации признаков (длина алфавита), m равно 26, количество признаков в словаре N равно 395.

Находим значение полинома $[d]$:

$$[d] = 97 \cdot 26^3 + 99 \cdot 26^2 + 97 \cdot 26^1 + 116 \cdot 26^0 = 1704872 + 66924 + 2522 + 116 = 17\,74\,434.$$

Тогда значение хеш-функции от слова «acat»

$$f(d) = ([d] \bmod N) + 1 = 1774434 \bmod (395) + 1 = 94 + 1 = 95.$$

К недостаткам метода поиска в словаре параметров, основанного на принципах хеш-функции, следует также отнести следующее требование: искажение символа в параметре поиска для хеш-словаря КА должно быть исключены полностью [8].

Анализ эффективности методов поиска признаков КА в словаре. Для обеспечения правдоподобности результатов в ходе исследования содержимое словаря было полностью поменяно десять раз, при этом количество элементов оставалось неизменным и составляло 395 слов. В результате для трех вышеуказанных методов поиска признаков КА в словаре были выявлены следующие показатели эффективности: «худший/лучший» варианты — соответственно

наибольшее/наименьшее количество итераций метода при поиске признаков на десяти различных словарях одинаковой емкости.

Значения эффективности методов поиска признаков КА в словаре представлены в табл. 2 и на рис. 3.

Таблице 2

Значения эффективности методов поиска признаков КА

Метод поиска	Количество итераций	
	Значение лучшего варианта	Значение худшего варианта
Последовательный	39	2495
Итерационный	399	402
На основе хеш-адресации	1 (при отсутствии коллизий)	

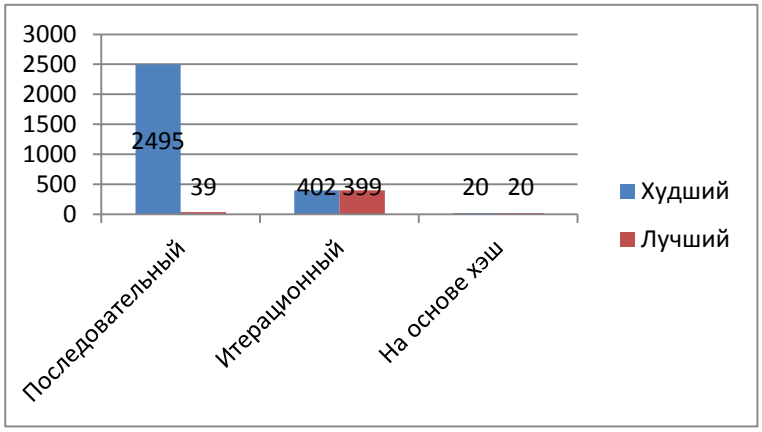

Рис. 3. Количество итераций лучшего/худшего вариантов для различных методов поиска признаков КА в словаре

Сводные результаты моделирования поиска для некоторых значений словаря приведены в табл. 3.

Таблица 3

Результаты моделирования (количество итераций)

№ п/п	Параметр поиска (слово)	Порядковый номер параметра в словаре	Метод поиска		
			Последовательный	Итерационный	На основе хеш-адресации
1	acat	1	39	399	1 (нет коллизий)
2	crisis	100	639	406	1 (нет коллизий)
3	iridium	200	1288	401	1 (нет коллизий)
4	property	300	1895	405	1 (нет коллизий)
5	whiteman	395	2495	402	1 (нет коллизий)

Количество итераций (временные затраты) для последовательного метода прямо пропорционально объему словаря и нахождению искомого параметра (слова) в нем. На рис. 4 представлено количество итераций поиска (вертикальная ось) в зависимости от места нахождения параметра в словаре (горизонтальная ось).

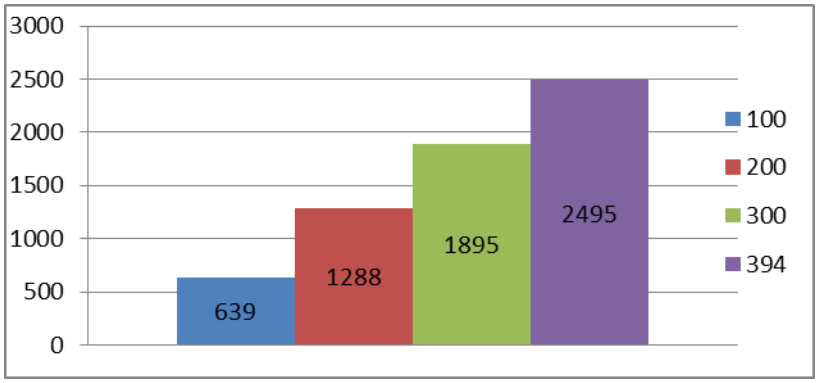

Рис. 4. Количество итераций для последовательного метода в зависимости от местонахождения параметра в словаре

Сравнение методов поиска признаков КА в словаре можно осуществлять по следующим показателям:

1) *сортировка результатов поиска*. Если результат поиска должен быть отсортирован, словарь признаков, основанный на хеш-таблице, представляется не вполне приемлемым, так как элементы словаря заносятся в таблицу в порядке, определяемом только их хеш-значениями. В случае последовательного и итерационного метода мы имеем отсортированный словарь;

2) *расходы на оперативную память ПЭВМ*. Для словаря признаков основанного на хеш-таблице требуется один указатель на значение и память под саму таблицу. Для последовательного и итерационного методов требуется выделять дополнительные ячейки памяти под таблицы;

3) *временные затраты* (или количество итераций) на поиск признаков в словаре различными методами. Наиболее эффективным методом поиска признаков КА при больших объемах данных и словаря является метод на основе использования хеш-функции. Для небольших объемов словаря допускается использовать последовательный или итерационный (предпочтительнее) методы;

4) *возможность осуществления поиска признаков КА в словаре при наличии ошибок/искажений в параметре поиска*. Наиболее устойчивым к искажениям символов в параметре поиска (при условии, что информация о степени искажения есть) является последовательный метод. При использовании метода поиска основе хеш-функции искажение символа в параметре поиска должны быть исключены полностью. Таким образом, при искажении отдельных символов параметра, только последовательный и итерационный методы позволяют принимать решение по результатам поиска [9].

В заключение сформируем краткие рекомендации специалистам по защите информации и обеспечения безопасности сетевых ресурсов по использованию трех рассмотренных разновидностей метода сигнатур:

1) если объем словаря признаков КА небольшой и поиск признаков КА осуществляется редко, допускается пользоваться последовательным методом для перебора всех значений словаря;

3) если объем словаря большой, и поиск признаков КА осуществляется часто, наиболее предпочтительно использование метода, основанного на хеш-адресации;

3) в случае относительно большой вероятности искажения символов в параметре поиска, а также наличия информации о степени искажения необходимо учитывать, что наиболее устойчивым к искажениям символов в параметре поиска является последовательный метод. Кроме того, допускается использование метода итераций. Наименее устойчив к искажением поискового слова метод на основе использования хеш-функции.

Литература

[1] Шаньгин В.Ф. *Защита информации в компьютерных системах и сетях*. Москва, ДМК Пресс, 2012, 592 с.

[2] Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушкин А.В. *Основы* криптографии. Москва, Гелиос АРВ, 2002, 480 с.

[3] Федоров А.К. Противодействие атаке “Photon number splitting attack” при квантовом распределении ключа BB84. *Студенческий научный вестник. Сб. тезисов общеуниверситетской науч.-тех. конф. «Студенческая научная весна-2011»*. Т. XI, ч. II. Москва, Изд-во МГТУ им. Н.Э. Баумана, 2012, с. 204–205.

[4] Абашев А.Н, Пазухин В.А, Слышкин А.С. На шаг впереди киберпреступников. *Информационная безопасность*, 2015, № 1, с. 8–11.

[5] Мазин А.В., Клочко О.С. Анализ методов противодействия угрозам и атакам на вычислительные системы. *Наукоемкие технологии в приборо- и машиностроении и развитие инновационной деятельности в вузе. Матер. Всеросс. науч. технич. конф. Т. 3*. 2014, с. 71–76.

[6] Климов С.М., Сычев М.П., Астрахов А.В. *Противодействие компьютерным атакам. Методические основы*. Москва, Изд-во МГТУ им. Н.Э. Баумана, 2013, 108 с.

[7] Фор А. *Восприятие и распознавание образов*. Москва, Машиностроение, 1989, 272 с.

[8] Харитонов В.А. *Основы теории живучести функционально избыточных систем*. Санкт-Петербург, Ин-т информатики и автоматизации РАН, 1993, 60 с.

[9] Гамаюнов Д.Ю. *Обнаружение компьютерных атак на основе анализа поведения сетевых объектов*. Дисс. … канд. физ.-мат. наук. Москва, МГУ, 2007, 88 с.

Череданова Екатерина Максимовна — студентка кафедры «Защита информации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Печкурова Ирина Руслановна — студентка кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Печкурова Виктория Руслановна — студентка кафедры «Специальная робототехника и мехатроника», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Мамченко Елизавета Андреевна — студентка кафедры «Защита информации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

COMPARATIVE ANALYSIS OF METHODS OF SEARCHING SIGNATURES OF COMPUTER ATTACKS IN THE DICTIONARY OF SYMBOLS

E.M. Cheredanova pankooova@mail.ru
SPIN-code: 1619-5499

I.R. Pechkurova risha.irisha18@mail.ru
SPIN-code: 2108-3379

V.R. Pechkurova vika.florida@gmail.com
SPIN-code: 4792-8460

E.A. Mamchenko liza.98.98@mail.ru
SPIN-code: 2887-3715

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

The article states that the ability to identify computer DDoS attacks in real time is one of the key aspects of information security, ensuring the security of network resources, and limiting access to server computing facilities. Analysis of the effectiveness of the three types (sequential, iterative and by using hash addressing) of the signature method based on the search for signs of computer attack in the corresponding dictionary is conducted. As a result of the research, the merits and demerits of all the methods considered were revealed. The factors influencing the decision-making on the choice of a particular type of signature method are also identified. In a generalized form, brief recommendations to specialists in protecting information and securing network resources on the use of the three variants of the signature method for detecting a DDoS attack are generated.

Keywords

Signature method, DDoS attack, feature dictionary, iteration, hash-function, sequential method, iterative method, hash addressing method

Received 28.05.2018

References

[1] Shan'gin V.F. Zashchita informatsii v komp'yuternykh sistemakh i setyakh [Information protection in PC networks and systems]. Moscow, DMK Press publ., 2012, 592 p.

[2] Alferov A.P., Zubov A.Yu., Kuz'min A.S., Cheremushkin A.V. Osnovy kriptografii [Cryptography fundamentals]. Moscow, Gelios ARV publ., 2002, 480 p.

[3] Fedorov A.K. Protivodeystvie atake "Photon number splitting attack" pri kvantovom raspredelenii klyucha BB84 ["Photon number splitting attack" resistance at quantum distribution of BB84 key]. *Studencheskiy nauchnyy vestnik. Sb. tezisov obshcheuniversitetskoy nauch.-tekh. konf. "Studencheskaya nauchnaya vesna — 2011". T. XI, ch. II* [Students science bulletin. Proc. Universiry Sci.-Tech. Conf "Students scientific spring-2011". Vol. XI. P. II]. Moscow, Bauman Press, 2012, pp. 204–205.

[4] Abashev A.N, Pazukhin V.A, Slyshkin A.S. One step ahead cybercriminals. *Informatsionnaya bezopasnost'* [Information Security], 2015, no. 1, pp. 8–11.

[5] Mazin A.V., Klochko O.S. Analiz metodov protivodeystviya ugrozam i atakam na vychislitel'nye sistemy [Analysis of counteraction methods to threats and attacks on computer

networks]. *Naukoemkie tekhnologii v priboro- i mashinostroenii i razvitie innovatsionnoy deyatel'nosti v vuze. Mater. Vseross. nauch. tekhnich. konf. T. 3* [High-end technologies in instrument and mechanical engineering and development of innovative activity in university. Proc. Russ. Sci.-Tec. Conf. Vol. 3]. Moscow, 2014, Bauman Press, pp. 71–76.

[6] Klimov S.M., Sychev M.P., Astrakhov A.V. Protivodeystvie komp'yuternym atakam. Metodicheskie osnovy [Counteractions to computer attacks. Methodical fundamentals]. Moscow, Bauman Press, 2013, 108 p.

[7] For A. Vospriyatie i raspoznavanie obrazov [Image acquisition and recognition]. Moscow, Mashinostroenie publ., 1989, 272 p.

[8] Kharitonov V.A. Osnovy teorii zhivuchesti funktsional'no izbytochnykh system [Theory fundamentals on survivability of functionally redundant systems]. Sankt-Petersburg, SPIIRAS publ., 1993, 60 p.

[9] Gamayunov D.Yu. Obnaruzhenie komp'yuternykh atak na osnove analiza povedeniya setevykh ob"ektov. Diss. kand. fiz.-mat. nauk [Computer attack detection based on analysis of network objects behavior. Kand. Phys.-Math. Sci. Diss.]. Moscow, MSU publ., 2007, 88 p.

Cheredanova E.M. — Bachelor's Degree student, Department of Information Security, Bauman Moscow State Technical University, Moscow, Russian Federation.

Pechkurova I.R. — Bachelor's Degree student, Department Automatic Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Pechkurova V.R. — Bachelor's Degree student, Department of Robotics and Mechatronics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Mamchenko E.A. — Bachelor's Degree student, Department of Information Security, Bauman Moscow State Technical University, Moscow, Russian Federation.