УДК 004.021

МНОГОУРОВНЕВЫЙ БИОНИЧЕСКИЙ АЛГОРИТМ ДЛЯ ОБНАРУЖЕНИЯ И ИДЕНТИФИКАЦИИ ПРОГРАММНО-АППАРАТНЫХ ВОЗДЕЙСТВИЙ НА ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ

Исупов Антон Борисович
Военная академия связи (филиал г. Краснодар), Краснодар, Россия

В статье описан разработанный многоуровневый бионический алгоритм обнаружения и идентификации программно-аппаратных воздействий на телекоммуникационные сети. Обсуждаются результаты применения программной реализации алгоритма

Ключевые слова: ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ, ПРОГРАММНО-АППАРАТНЫЕ ВОЗДЕЙСТВИЯ, АЛГОРИТМ, БИОНИКА, МНОГОУРОВНЕВЫЙ ПОДХОД

UDC 004.021

MULTILEVEL BIONIC ALGORITHM FOR DETECTION AND IDENTIFICATION OF SOFTWARE AND HARDWARE IMPACTS ON TELECOMMUNICATION NETWORKS

Isupov Anton Borisovitch
Military academy of communication corps (Krasnodar branch office), Krasnodar, Russia

Multilevel bionic algorithm for detection and identification of software and hardware impacts on telecommunication networks is considered in the article. The results of applying of the software implementation of the algorithm are discussed

Keywords: TELECOMMUNICATION NETWORKS, SOFTWARE AND HARDWARE IMPACTS, ALGORITHM, BIONICS, MULTILEVEL TREATMENT

Введение

Осуществление злоумышленниками все более изощренных программно-аппаратных воздействий (ПАВ) в информационно-телекоммуникационных сетях (ИТС) [1] показывает, что уязвимости протокола передачи данных и операционных систем, потенциальные ошибки в специальном программном обеспечении [2, 3] обусловили необходимость создания самостоятельных средств обнаружения и идентификации программно-аппаратных воздействий на ИТС. Современные средства защиты информации от НСД и средства защиты удаленного доступа к информации не обеспечивают требуемый уровень обнаружения и идентификации программно-аппаратных воздействий.

В настоящее время известно более 20 средств обнаружения и идентификации программно-аппаратных воздействий, основанных на сигнатурных и поведенческих методах анализа, а также комбинаций этих методов [4].

Исследование, проведенное компанией NNS Group в 2011 году [5], выявило усредненные характеристики существующих в настоящее время средств обнаружения и идентификации программно-аппаратных воздействий (табл. 1).

Таблица 1 – Характеристики средств обнаружения и идентификации ПАВ

Характеристика	**Значение**
Вероятность выявления известных типов ПАВ ($P_д^{И}$)	0,95
Вероятность выявления неизвестных типов ПАВ ($P_д^{Н}$)	0,7
Количество ошибок первого рода (чужой как свой)	18 %
Количество ошибок второго рода (свой как чужой)	не менее 5 %
Уровень загрузки аппаратных ресурсов	до 48 %

К числу основных недостатков современных средств обнаружения и идентификации программно-аппаратных воздействий можно отнести:

- необходимость обновления баз сигнатур через разработчика средства;

- отсутствие комплексного подхода к процессу обнаружения и идентификации ПАВ;

- низкую вероятность обнаружения и идентификации неизвестных типов ПАВ;

- большое количество ошибок первого рода;

- высокую вероятность ложных обнаружений;

- высокие требования к аппаратным ресурсам ИТС.

Все вышеизложенное обусловило необходимость поиска решения важной **научной задачи** – разработки современного средства обнаружения и идентификации ПАВ на ИТС, функционирующего на основе предложенного автором алгоритма, представляющего собой не комбинацию, а гибрид существующих алгоритмов сигнатурного и

поведенческого анализа и позволяющего обеспечить требуемую вероятность обнаружения как известных, так и неизвестных типов программно-аппаратных воздействий на ИТС в реальном масштабе времени с минимально возможным количеством ошибок первого и второго рода.

В рамках решения поставленной научной задачи был разработан многоуровневый бионический алгоритм обнаружения и идентификации программно-аппаратных воздействий на ИТС, сочетающий в себе преимущества сигнатурного и поведенческого анализа и функционирующий по принципу иммунной системы человека.

Описание алгоритма

Многоуровневый бионический алгоритм обнаружения и идентификации программно-аппаратных воздействий на ИТС основывается на разделении процесса обнаружения и идентификации ПАВ в три этапа применительно к трем основным этапам реализации программно-аппаратного воздействия и механизмам искусственных иммунных систем [6, 7, 9] в сочетании с репродуктивным планом Холланда (РПХ) для обеспечения требуемой скорости и качества процесса обнаружения.

В обобщенном виде работа многоуровневого бионического алгоритма обнаружения и идентификации программно-аппаратных воздействий на ИТС показана на рисунке 1.

Алгоритм реализует два режима работы:

- ***режим обучения***;

- ***оперативный режим***.

При работе алгоритма в ***режиме обучения*** происходит формирование образа нормальной активности объекта ИТС (ОНА ИТС),

нормальных шаблонов активности и множества детекторов, являющихся своего рода сигнатурами программно-аппаратных воздействий.

Для описания работы алгоритма вводятся следующие обозначения:

L – строка кандидатов в детекторы;

S_L – набор кандидатов в детекторы;

N_{RO} – число строк-кандидатов в детекторы, сформированных алгоритмом генерации;

R – детектор;

N_R – число детекторов, оставшихся после удаления, совпавших с нормальными шаблонами активности;

W – нормальный шаблон активности;

N_w – число нормальных шаблонов активности;

P_m – вероятность совпадения двух сгенерированных строк;

$f = (1 - P_m)_{N_S}$ – вероятность того, что сгенерированная строка не совпадет ни с одним из N_S нормальных шаблонов активности;

P_{f_n} – вероятность того, что оставшиеся N_R детекторов не смогут обнаружить ПАВ;

$P_{o_n} = 1 - P_{f_n}$ – вероятность обнаружения программно-аппаратного воздействия на одном из этапов обнаружения.

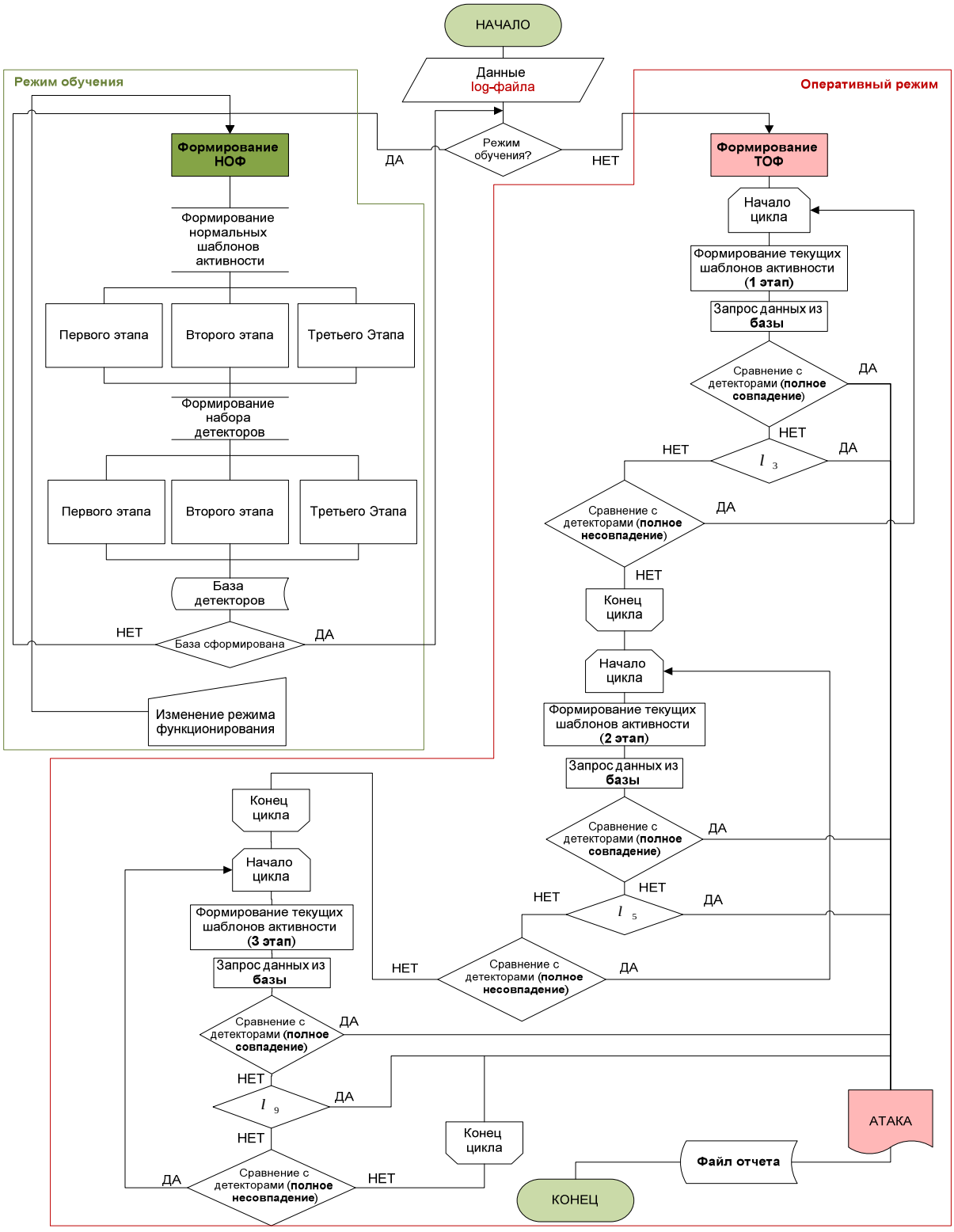

Рисунок 1. Структурная схема многоуровневого бионического алгоритма обнаружения и идентификации ПАВ на ИТС

Будем полагать, что входными данными для формирования образа нормальной активности объекта ИТС являются данные с модулей сбора

сетевых данных, модулей регистрации узловых событий, а также записи log-файлов объекта ИТС. Эти данные представляют собой текстовые записи, которые отражаются в log-файле системы в порядке их появления. Таким образом, в log-файле системы обнаружения будут отражены все события, произошедшие на защищаемом объекте за определенный промежуток времени, достаточный для описания состояния его нормального функционирования. Каждой записи в log-файле присваивается 32-битный идентификационный код.

По истечении заданного интервала времени данные, собранные для объекта ИТС, обрабатываются следующим образом:

1) Выбирается цепочка событий (записей в log-файле) – изменение в последовательности, которая может указывать на проведение определенного этапа программно-аппаратного воздействия на ИТС (рис. 2).

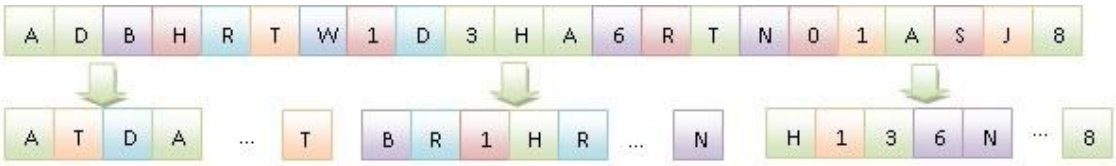

Рисунок 2. Процесс формирования ОНА ИТС

2) Выбранная цепочка событий представляется в виде ряда дискретных элементов, составленных из букв конечного алфавита.

3) Выбирается размер окна *l*, определяющего длину нормального шаблона активности.

4) Окно движется вдоль ряда данных с шагом смещения, определяемым параметром *Q*.

Размер окна *l* и шаг смещения *Q* выбирается таким образом, чтобы обеспечить требуемый порог чувствительности на каждом этапе работы алгоритма.

5) Выделенные данные сохраняются в базе как шаблоны нормальной активности. Все определенные шаблоны составляют часть образа нормальной активности объекта ИТС (рис. 3).

По такому же принципу формируются шаблоны нормальной активности второй и третьей частей ОНА ИТС.

Количество шаблонов нормальной активности N_W каждой из частей нормального образа функционирования объекта ИТС зависит от времени *T*, отведенного для характеристики состояния его нормального функционирования и является конечным множеством.

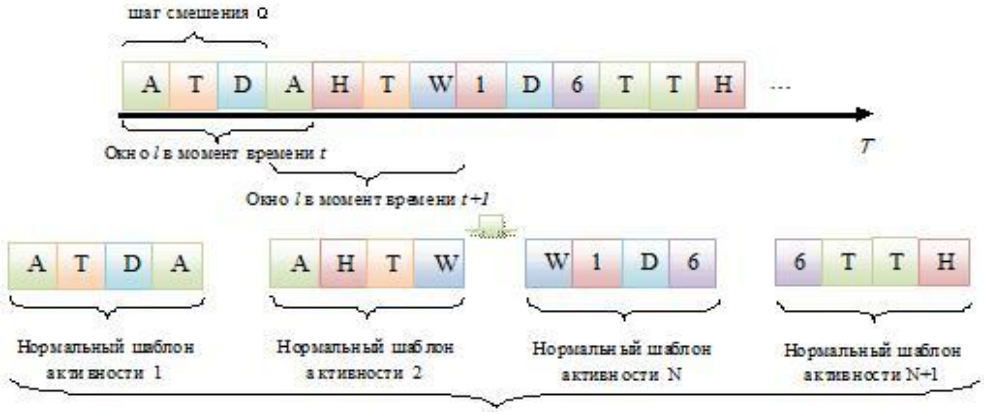

Рисунок 3. Процесс формирования нормальных шаблонов активности

Далее происходит процесс создания и эволюции генной библиотеки (базы детекторов).

На первом этапе работы алгоритма генерации, в качестве *R* используются шаблоны нормальной активности W части ОНА объекта ИТС. В последующем, генерация детекторов происходит с применением РПХ.

Основными понятиями репродуктивного плана Холланда являются:

- *ген* – атомарный элемент хромосомы;

- *аллель* – значение конкретного гена;

- *локус* – позиция гена в хромосоме;

- *хромосома* – упорядоченная последовательность генов;

- *популяция* – конечное множество хромосом.

При использовании репродуктивного плана Холланда совместно с методами искусственных иммунных систем можно совместить понятия *хромосома* и *детектор*, *популяция* и *набор детекторов*, *ген* и *символ конечного алфавита*.

Каждый новый набор кандидатов в детекторы S_L выбирается из полученных хромосом при очередном использовании РПХ.

Каждое использование РПХ состоит из следующих этапов:

1) Инициализация.

На этапе инициализации выбирается исходная популяция хромосом, состоящая из хромосом, полученных в результате предыдущего использования РПХ (рис. 4).

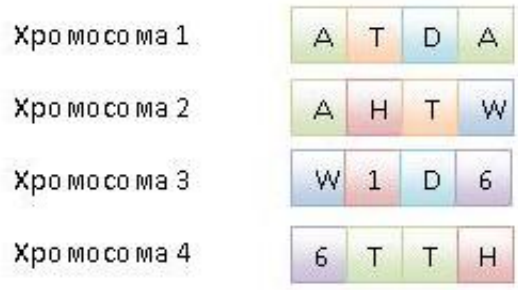

Рисунок 4. Инициализация РПХ

2) Оценка приспособленности хромосом в популяции.

Оценка приспособленности хромосом состоит в расчете функции приспособленности F_{pr} для каждой хромосомы в популяции [6]:

$$F_{pr} = 1 - \frac{r}{l}, \quad (1)$$

где r – максимальное количество генов в хромосоме, совпавших с генами одной из исходных хромосом; l – длина хромосомы (рис. 5).

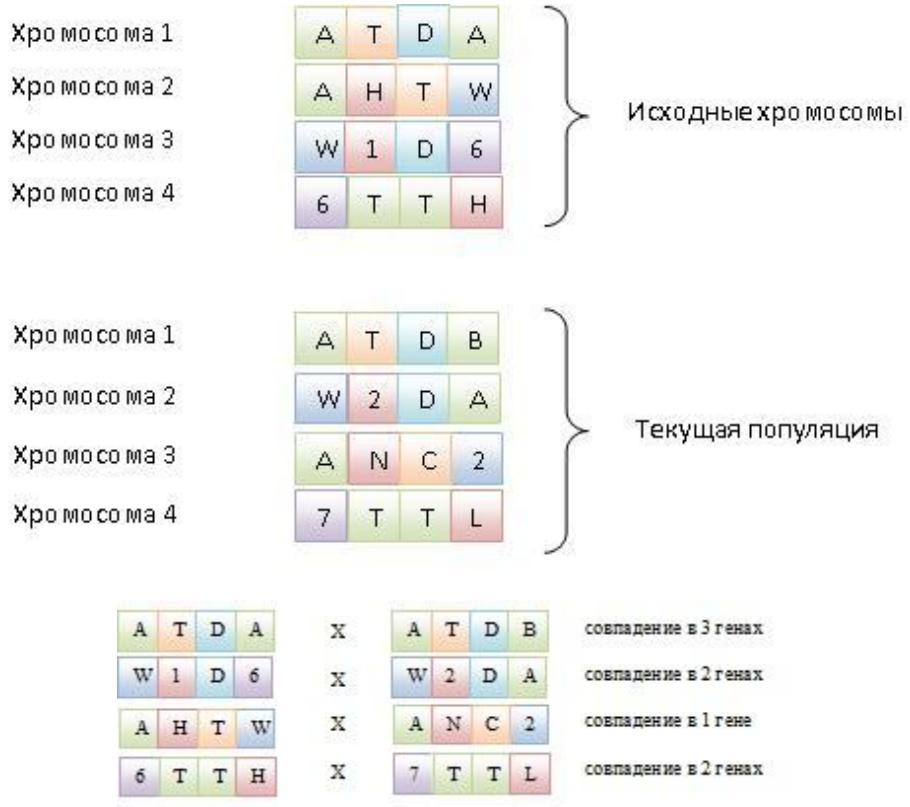

Рисунок 5. Оценка приспособленности хромосом в популяции

Значение функции приспособленности для рассматриваемой популяции приведено в таблице 2.

Таблица 2 – Значение функции приспособленности

Хромосома	Длина хромосомы (l)	Количество совпавших генов (r)	Значение функции приспособленности (F_{pr})
1	4	3	0,25
2	4	2	0,5
3	4	1	0,75
4	4	2	0,5

3) *Проверка условия окончания работы РПХ.*

Условием окончания работы РПХ является получение популяции хромосом N_R, каждая из которых не совпадает ни с одной из хромосом множества имеющихся кандидатов в детекторы S_L.

В случае, если условие окончания работы РПХ выполняется, текущая популяция хромосом N_R добавляется к набору кандидатов в детекторы S_L, в противном случае, происходит переход на следующий шаг РПХ.

4) *Селекция хромосом*.

Селекция хромосом заключается в выборе тех хромосом, которые будут участвовать в создании потомков для следующей популяции. Этот выбор происходит в соответствии с принципами естественного отбора, когда наибольшие шансы быть выбранными в родители имеют хромосомы с наибольшими значениями функции приспособленности.

Для выбора родительских пар хромосом используется метод Монте-Карло (рис. 6). Каждой хромосоме $X_1, X_2, \ldots, X_n, \quad n = 1,2,\ldots,N$ (где N – численность популяции) соответствует сектор колеса рулетки v_{Xn}, выраженный в процентах [6]:

$$v_{x_n} = p(X_n) \times 100\%. \qquad (2)$$

где

$$p(X_n) = \frac{F_{pr}(X_n)}{\sum_{n=1}^{N} F_{pr}(X_n)}. \qquad (3)$$

Здесь $F_{pr}(X_n)$ – значение функции приспособленности хромосомы X_n; $p(X_n)$ – вероятность селекции хромосомы X_n. Чем больше сектор колеса, тем больше вероятность выбора соответствующей хромосомы. Выбор хромосомы осуществляется случайным выбором числа из интервала [0,100].

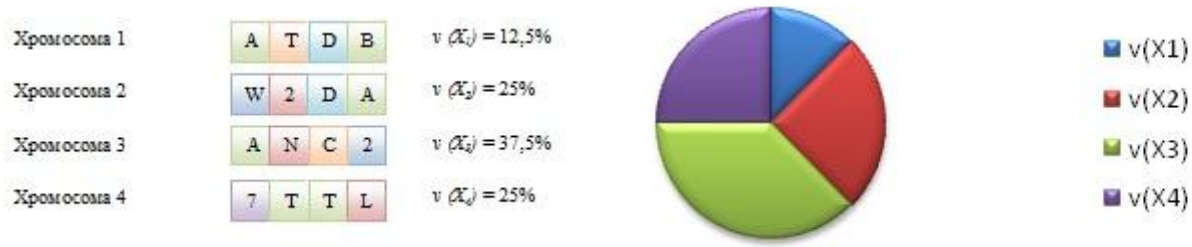

Рисунок 6. Выбор хромосом методом Монте-Карло

В результате этого процесса создается родительская популяция с численностью, равной численности текущей популяции

5) Применение генетических операторов.

Данный этап приводит к образованию новой популяции хромосом (потомков) от родительской популяции. Применяются два основных оператора: *кроссинговер* и *оператор мутации*.

На первом этапе скрещивания выбираются пары хромосом из родительской популяции. Эта популяция состоит из хромосом, отобранных на этапе селекции. Объединение хромосом в пары производится случайным образом с вероятностью скрещивания P_s. Далее для упрощения работы предлагается использовать вероятность скрещивания $P_s = 0{,}5$.

Для каждой пары отобранных родителей выбирается точка скрещивания (локус) – позиция, в которой каждый из родителей будет разделен на 2 части. Выбор точки скрещивания сводится к случайному выбору числа k из интервала [1, l-1]. В результате работы оператора кроссинговера получается пара потомков (рис. 7).

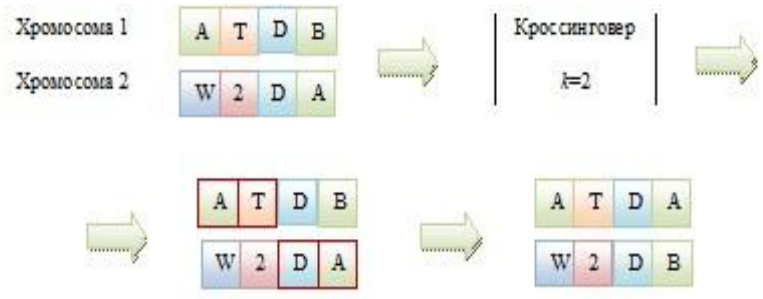

Рисунок 7. Процесс кроссинговера

Оператор мутации изменяет значение гена в хромосоме на иное с вероятностью P_m (рис. 8). Далее для упрощения работы предлагается использовать вероятность скрещивания $P_m = 0{,}01$.

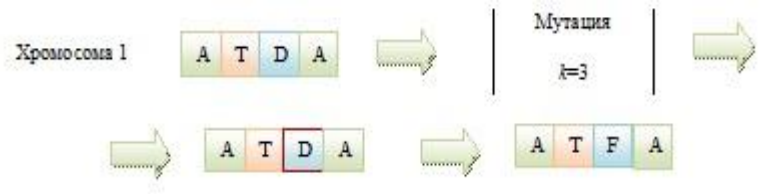

Рисунок 8. Процесс мутации

6) Формирование новой популяции.

Хромосомы, полученные в результате выполнения генетических операторов над родительской популяцией, включаются в состав новой популяции для очередной итерации РПХ.

При каждой итерации РПХ рассчитывается значение функции приспособленности F_{pr} для каждой из хромосом этой популяции, после чего проверяется условие окончания работы РПХ.

Процесс создания детекторов из шаблонов нормальной активности 2 и 3 частей ОНА аналогичен процессу создания детекторов 1-й части ОНА за исключением того, что для создания самих шаблонов нормальной активности используются другие временные цепочки событий, изменение в которых может указывать на 2 и 3 этапы реализации программно-аппаратного воздействия на ИТС.

На практике элементы множества детекторов создаются последовательно и продолжают генерироваться до тех пор, пока не будет достигнуто необходимое число детекторов [7].

Количество детекторов, необходимых для решения задачи обнаружения и идентификации программно-аппаратного воздействия с требуемым уровнем вероятности определяется следующим образом.

Еесли P_m достаточно мала, а N_W и N_R достаточно велики, то можно записать [7]:

$$f = (1 - P_m)^{N_W}. \quad (4)$$

$$N_R = N_{RO} \cdot f. \quad (5)$$

$$P_{fn} = (1 - P_m)^{N_w} \approx e^{-P_m N_R}. \quad (6)$$

Отсюда получаем:

$$N_R = -\frac{lnP_{fn}}{P_m}. \quad (7)$$

и, соответственно:

$$N_{RO} = \frac{N_R}{f} = -\frac{lnP_{fn}}{P_m (1 - P_m)^{N_w}}. \quad (8)$$

Формула (8) определяет число строк кандидатов в детекторы, которые необходимо сгенерировать на этапе обучения, в функции от вероятности необнаружения ПАВ P_{fn}, числа нормальных шаблонов активности N_w, и вероятности совпадения двух сгенерированных строк P_m.

$$P_m \approx m^{-r}\left[\frac{(l-r)(m-1)}{m} + 1\right], \quad (9)$$

где m – количество символов алфавита, r – максимальное количество генов в хромосоме, совпавших с генами одной из исходных хромосом; l – длина хромосомы.

При завершении процесса создания и эволюции генной библиотеки, т.е. когда завершается процесс формирования множества детекторов, алгоритм переходит из функционирования в «режиме обучения» в «оперативный режим».

При работе алгоритма в «оперативном режиме» происходят формирование образа текущей активности объекта ИТС (ОТА ИТС), формирование шаблонов текущей активности (ШТА) и их поэтапное сравнение с множеством сформированных в «режиме обучения» детекторов.

Входными данными для формирования ОТА ИТС, как и для ОНА ИТС, являются данные log-файла системы обнаружения.

Этапы формирования ОТА и ШТА ИТС

1) На данном этапе выбирается цепочка событий (записей в log-файле) – изменение в последовательности, которой указывает на проведение 1 этапа ПАВ на ИТС (рис. 9).

2) Выбранная цепочка событий (записей в log-файле) представляется в виде ряда дискретных элементов, составленных из букв конечного алфавита.

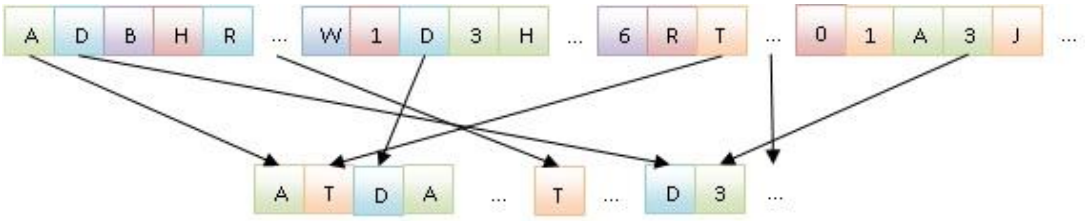

Рисунок 9. Процесс формирования ТИП

3) Выбирается размер окна *l*, равный длине нормального шаблона активности.

4) Окно движется вдоль ряда данных с шагом смещения *Q*, определенным для процесса формирования нормальных шаблонов активности.

5) Все выделенные шаблоны образуют множество шаблонов текущей активности одной из трех частей ОТА ИТС и сравниваются на совпадение с базой детекторов, образованной на основе шаблонов нормальной активности.

Сравнение шаблонов текущей активности с базой детекторов происходит на каждом уровне в три этапа. Сначала происходит сравнение очередного текущего шаблона активности на полное соответствие детекторам из базы. Если на первом этапе сравнения не находится однозначного соответствия шаблона текущей активности детектору в базе, то он проверяется на совпадение с базой детекторов по λ, определенной для данного этапа. На последнем этапе шаблон текущей активности проверяется на полное несовпадение с детекторами базы. Сравнение

текущих шаблонов активности с детекторами из базы осуществляется при помощи лингвистического подхода к распознаванию образов [8].

Любое совпадение шаблонов текущей активности с детекторами любого уровня из множества базы детекторов означает, что на ИТС осуществляется программно-аппаратное воздействие.

С целью оценки качества работы была разработана программная реализация вышеизложенного алгоритма и проведен эксперимент по обнаружению и идентификации ПАВ на ИТС.

В ходе проведения эксперимента по обнаружению и идентификации программно-аппаратных воздействий на ИТС в условиях инструментального стенда получены следующие результаты (табл. 3).

Таблица 3 – Результаты стендовых испытаний

№ п/п	Название ПАВ	Кол-во тестов	Ресурс	Тип	% обнаруженных ПАВ	Кол-во сообщений о ПАВ	Ошибки 1-го рода	Ошибки 2-го рода
1.	Proftp_put_down1	126	ftp	rr	95	120	6	2
2.	Proftp_put_down2	126	ftp	rr	97	122	4	2
3.	propro	40	ftp	rr	98	39	1	0
4.	babcia	40	ftp	rr	95	38	2	0
5.	Squ1RT	56	http	dos	95	53	3	0
6.	ПАВ 1	56	http	rr	96	54	2	1
7.	ПАВ 2	56	ssh	rr	95	53	3	1
8.	ПАВ 3	30	dcom	rr	93	28	2	0
9.	ПАВ 4	30	dcom	rr	86	26	4	0
10	ПАВ 5	100	0-1024	ps	97	97	3	1

Число сообщений о ПАВ во всех тестах не превышает число проведённых программно-аппаратных воздействий, поскольку система, функционирующая по вышеописанному алгоритму, была сконфигурирована таким образом, чтобы избежать повторяющихся

сообщений о ПАВ, что в общем случае понижало бы информативность системы для оператора, хотя и не влияло бы на качество обнаружения и идентификации ПАВ.

Вероятность обнаружения составила:

- **общая вероятность обнаружения** – 0,9984;
- **вероятность обнаружения неизвестных типов ПАВ** – 0,95;
- **вероятность обнаружения известных типов ПАВ** – 0,96.

Доля ложных срабатываний при обнаружении и идентификации ПАВ на ИТС составила 1,0 %, что в 6 раз ниже, чем аналогичный показатель у существующих на настоящий момент средств обнаружения, а доля необнаруженных ПАВ не превысила значения в 5,2 %.

В описанных тестах загрузка центральных процессоров ПЭВМ с установленной системой, на которые приходилась максимальная вычислительная нагрузка, не превышала 26 %.

Выводы. Разработанный и описанный в статье многоуровневый бионический алгоритм обнаружения и идентификации программно-аппаратных воздействий на ИТС позволил создать экспериментальное программное средство обнаружения, способное в реальном масштабе времени, не превышая временные показатели обнаружения существующих средств, повысить на 0,25 вероятность обнаружения как известных, так и неизвестных типов ПАВ на ИТС, что подтверждают приведенные результаты эксперимента.

Список литературы

1. Исупов А.Б., Юрков В.А., Королев И.Д., Скуратов В.Ю. Роль и место процесса обнаружения и идентификации программно-аппаратных воздействий в обеспечении устойчивого функционирования автоматизированных систем военного назначения // Информационная безопасность – актуальная проблема современности: Материалы III научно-технической школы-семинары / Военная академия вязи (филиал г. Краснодар). – Краснодар, 2011. – С. 6–11.
2. Белый А.Ф. Сетецентризм и стратегия непрямых действий // Материалы научно-практической конференции «Информационная безопасность автоматизированных систем управления военного назначения-2009» /

Ростовский военный институт Ракетных войск. – Ростов-на-Дону, 2009. – С. 20–28.
3. Климов С.М. Методы и модели противодействия компьютерным атакам. – Люберцы.: КАТАЛИТ, 2008. – 316 с.
4. Исупов А.Б., Юрков В.А., Королев И.Д., Скуратов В.Ю. Средства обнаружения и идентификации программно-аппаратных воздействий на АС ВН // Информационная безопасность – актуальная проблема современности: Материалы III научно-технической школы-семинары / Военная академия связи (филиал г. Краснодар). – Краснодар, 2011. – С. 12–16.
5. Отчет компании NNS Group «NNS Group IDS GroupTest Report (edition 1)», 2011. – 68 с.
6. Исупов А.Б., Королев И.Д., Юрков В.А., Скуратов В.Ю. К вопросу об использовании механизмов искусственных иммунных систем для решения задач обнаружения и идентификации неизвестных типов программно-аппаратных воздействий на автоматизированные системы // Информатика: проблемы, методология, технологии: Материалы XII международной конференции; Воронежский государственный университет. – Воронеж, 2012.
7. Васильев В.И. Интеллектуальные системы защиты информации: учеб. Пособие. – М.: Машиностроение, 2010. – 152 с.
8. Дж. Вэн Райзин. Классификация и кластер. – М.: Мир, 1980. – 389 с.
9. Д. Дасгупта. Искусственные иммунные системы и их применение. – М.: Физматлит, 2006. – 344 с.