УДК 004.853

05.00.00 Технические науки

МЕТОДИКА ОБНАРУЖЕНИЯ ПОЛИМОРФНЫХ ВИРУСОВ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ И ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Частикова Вера Аркадьевна
к.т.н., доцент
РИНЦ SPIN-код: 4525-0290
chastikova_va@mail.ru

Берёзов Максим Юрьевич
студент
maxberezov@gmail.com
Кубанский государственный технологический университет, Краснодар, Россия

Настоящая статья посвящена исследованию фундаментальных свойств и компонент иммунной системы, таких как, В-лимфоциты, Т-лимфоциты, иммунологическая память системы, первичный и вторичный иммунный ответ, иммунологическое обучение детекторов, которые составляют базис разработанной методики обнаружения полиморфных вирусов. Ключевую роль в полученной методике играют объекты-детекторы, которые сочетают в себе свойства В- и Т-лимфоцитов, участвующих в реакциях врожденного и адаптивного иммунного ответа. Полиморфизм компьютерных вирусов заключается в формировании кода вредоносной программы прямо во время ее исполнения, причем функция, отвечающая за формирования кода, не является статической. Таким образом, не удается однозначно создать сигнатуру, соответствующую данному полиморфному вирусу. В статье выделены и описаны основные функции искусственных иммунных систем, предложены идеи их реализации, а также программного, системного взаимодействия. Построена таблица, отражающая соответствие между компонентами искусственной иммунной системы и иммунной системой позвоночных. У иммунных систем выявлены такие важные особенности, реализация которых будет эффективна при решении задач обнаружения зловредного программного кода. Для более продуктивного обучения системы рассматривается класс генетических операторов, таких как, кроссинговер, мутация, селекция, описывается их абстрактная реализация. Каждый оператор реализован в нескольких вариантах, рассматривается совокупность подходов для их имплементации в виде конкретной системы. Построена система взаимодействия генетических и иммунологических алгоритмов

Ключевые слова: ИСКУССТВЕННАЯ

UDC 004.853

Technical sciences

METHOD OF POLYMORPHIC VIRUSES DETECTION BASED ON ARTIFICIAL IMMUNE SYSTEM AND GENETIC ALGORITHMS

Chastikova Vera Arkadyevna
Cand.Tech.Sci., associate professor
RSCI SPIN-code: 4525-0290
chastikova_va@mail.ru

Berezov Maxim Yurievich
student
maxberezov@gmail.com
Kuban State Technological University, Krasnodar, Russia

This article is dedicated to the study of the fundamental properties and components of the immune system such as B lymphocytes, the T-lymphocytes, immune system storage, primary and secondary immune response, immunological training detectors, which will be the basis of the obtained as a result of detection methods of polymorphic viruses using artificial immune systems. Polymorphism of computer viruses is the formation of a malicious program code directly during execution. Thus, it is impossible to create a unique signature corresponding to these polymorphic viruses. A similar classification problem is solved by the immune system of vertebrates, stared again met with the virus, it "remembers" him, and the next time provides effective secondary immune response. These properties of the immune system served as a prerequisite for the use of immune approaches and algorithms for solving the problems of detection of malicious code. The article identified and described their main features, proposed the idea of their implementation and software, system interactions in the immune system revealed such important features, the implementation of which will be effective in solving the problem of detection of malicious code and software. Also, for a more productive system of education is considered a class of genetic, evolutionary algorithms, described by their immediate implementation of site-specific decentralized artificial immune system, built a system of interaction of genetic and immunological algorithms.

Keywords: ARTIFICIAL IMMUNE SYSTEM,

ИММУННАЯ СИСТЕМА, ЭВРИСТИЧЕСКИЙ АЛГОРИТМ, ОБНАРУЖЕНИЕ ВИРУСОВ, БИОИНФОРМАТИКА, ГЕНЕТИЧЕСКИЙ АЛГОРИТМ

HEURISTIC ALGORITHM, VIRUS DETECTION, BIOINFORMATICS, GENETIC ALGORITHM

Doi: 10.21515/1990-4665-124-048

Многие концепции и алгоритмы в информатике были подсказаны самой природой. В частности, иммунная система человека классифицирует с большой точностью все поступающие в нее клетки на свои и чужеродные. Данный факт является предпосылкой использования принципов и свойств иммунной системы в задачах обнаружения полиморфных вирусов. Кроме того, все жизненно важные системы человека находятся в непрерывном эволюционном развитии, что говорит о возможности применения класса генетических алгоритмов для более быстрого обучения и развития объектов системы.

1 Проблематика исследования

Для моделирования искусственной иммунной системы (ИИС) необходимо выделить главные компоненты иммунной системы (ИС) человека, определить основные свойства, которые будут реализованы, и составить математическую модель ее функционирования.

Вся классификация в иммунной системе человека на «свои» и «чужие» клетки основана на химических связях, которые образуются между белковыми цепями. В работе смоделированы белковые цепи в виде HEX–строк определенной длины. Множество всех возможных строк делится на два подмножества: «свои» и «чужие». Перед искусственной иммунной системой стоит задача классификации строки, то есть отнесения ее к одному из двух подмножеств. В подобной модели теоретически возможны два вида ошибок – «чужая» строка классифицируется как «своя», и «своя» строка классифицируется как «чужая». [2] Целью разработанной ИИС является снижение уровня этих ошибок. Данная модель применима в задачах обнаружения вирусов. [7,8]

2 Иммунная система

Иммунная система состоит из множества клеток и молекул, которые взаимодействуют между собой различными способами для обнаружения и уничтожения инфекционных патогенов (болезнетворных микроорганизмов). Поверхности клеток иммунной системы покрыты рецепторами, некоторые из которых химически связаны с патогенами, некоторые - с другими клетками ИС: все это необходимо для построения сложной модели сигнализации, которая вызовет иммунный ответ. Большинство клеток ИС циркулируют по всему организму через кровь и лимфу, образуя систему распределенного обнаружения и реагирования, где нет централизованного управления. Обнаружение и устранение патогенов является результатом работы триллионов клеток, взаимодействующих с помощью простых, локализованных правил. Следствием этого является то, что ИС очень устойчива к выходу из строя отдельных компонентов и атак на саму систему. [7,9]

2.1 Лимфоциты (детекторы)

ИС млекопитающих состоит из множества видов клеток, которые могут выступать в роли детектора. В работе использован базовый класс детекторов, смоделированный на основе свойств лимфоцитов. Этот класс включает свойства В – лимфоцитов, Т – лимфоцитов и антител.

Лимфоциты на своей поверхности содержат множество рецепторов, данные рецепторы связаны с определенной частью патогена, которая определяется его химической структурой, так что лимфоцит может помечать в качестве чужеродного объекта тот патоген, на который среагировал его рецептор. Подобная схема реализована в работе.

Каждый рецептор и патоген представлен HEX – строкой фиксированной длины, химическая связь между ними смоделирована на основе соответствия строки рецептора строке патогена. Приведем два правила, показывающие степень соответствия двух строк.

1) Расстояние Левенштейна – это минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую. Расстояние Левенштейна для бинарных строк определятся как сумма числа позиций, в которых символы строк различны.

2) Подсчет значения r – максимального числа смежных бит. Данный подход нашел широкое применение в биологических исследованиях, суть его заключается в нахождении максимальной одинаковой подстроки. В работе r задано таким образом, чтобы совпадающая подстрока была длиной не более r.

Активация лимфоцита произойдет тогда и только тогда, когда заранее заданное количество рецепторов среагируют на патоген.

2.2 Подготовка системы обнаружения, иммунологическое обучение клеток

Лимфоциты называют отрицательными детекторами, так как они создаются с целью реагировать на чужеродные клетки. Когда лимфоцит переходит в активное состояние, иммунная система понимает, что был обнаружен чужеродный объект. Эта форма обучения называется толерантностью, так как лимфоциты подготавливаются, чтобы быть толерантными по отношению к себе.

Лимфоциты на начальном этапе состоят из случайно сгенерированных рецепторов, поэтому могут реагировать и на чужие, и на свои клетки. Рассмотрим один из классов лимфоцитов – Т-лимфоциты. Они образуют в тимусе, это единственное место, где они являются толерантными; там же незрелая Т–клетка развивается, и если в процессе развития она приходит в активное состояние, то умирает. Почти все белковые клетки организма экспессируются в тимусе, поэтому вышедшие оттуда Т– клетки будут толерантны по отношению к другим белковым клеткам организма. Этот процесс называется отрицательной селекцией.

2.3 Память иммунной системы

Иммунная система обладает адаптивным ответом, который позволяет определить структуру патогена, определить, опасен ли он и «запомнить» его, чтобы дать еще более быстрый и точный ответ при будущей встрече с ним. Таким образом, ИС «помнит» те патогены, на которые среагировали ее лимфоциты. Различают первичный и вторичный ответ; при первой встрече лимфоцита с патогеном иммунная система человека уничтожает и запоминает патоген, как правило, в течение нескольких недель. При второй и последующих встречах ИС дает более эффективный и быстрый ответ.

Первичный ответ достаточно медлителен, так как малое количество лимфоцитов связано с новым вирусом, поэтому активные лимфоциты начинают клонировать себя, наблюдается экспоненциальный рост их популяции - чем больше соответствие между рецепторами лимфоцита и патогеном, тем больше вероятность, что лимфоцит будет клонирован.

Заметим, что со временем патогены также клонируются, происходит гонка между ними и лимфоцитами. Шансы иммунной системы увеличиваются за счет так называемых В–лимфоцитов, которые подвергаются мутации в процессе клонирования. После уничтожения инфекции ИС определяет лимфоциты с наибольшей мерой аффинности и стабилизирует размер популяции, убивая мало приспособленные лимфоциты. Размер полученной популяции будет достаточен, чтобы дать быстрый и эффективный ответ на новый патоген. Важным свойством является то, что у лимфоцитов с большой мерой аффинности (мерой соответствия наибольшему количеству патогенов) снижен порог активации. Данные свойства ИС реализованы в программном комплексе, благодаря им значительно увеличивается скорость вторичного ответа.

3 Генетические алгоритмы

Для повышения эффективности работы искусственной иммунной системы в работе предлагается использовать класс генетических алгоритмов, которые представляют собой эвристические методы оптимизации, основанные на определении лучших представителей своей популяции, наиболее приспособленных к текущим условиям, и передачи генов будущим потомкам. На первые роли выходят такие операторы как скрещивание, селекция, мутация и кроссинговер. [1,3-6]

3.1 Кроссинговер

Под кроссинговером понимают обмен участками гомологичных хромосом. Существует два подхода к реализации этого оператора:

- создание случайной маски;
- склеивание родительских хромосом относительно точки (одноточечный кроссинговер).

Оператор кроссинговера (создание случайной маски) применяется к двум детекторам-родителям (обозначим их родитель1 и родитель2), на выходе получается новый детектор-потомок. На первом этапе кроссинговера для нового детектора составляют маску длиной, равной количеству бит, которые потомок должен унаследовать от родителей. Маску заполняют случайными значениями: 1 или 0. Если значение маски равно 1, то данный бит унаследуется от первого родителя, иначе – от второго. В табл. 1 приведен пример набора родительских бит и бит потомка.

Таблица 1 – Набор родительских бит и бит потомка

Родитель1	1	1	1	1	1	1	1	1
Родитель2	0	0	0	1	0	1	0	0
Маска	1	0	1	0	1	0	1	0
Потомок	1	0	1	1	1	1	1	0

3.2 Мутация

Для увеличения структурного разнообразия детекторов введен

оператор мутации. Мутация может осуществляться посредством нескольких алгоритмов мутации бинарной строки:

– одноточечный оператор мутации;

– многоточечный оператор мутации;

– инверсия k – случайных бит;

– инверсия бит относительно выбранной точки разреза.

Одноточечный оператор мутации заключается в выборе произвольной точки строки, и перестановки ее значения с соседней точкой.

Многоточечный оператор мутации заключается в выборе произвольной точки строки, которой будет присвоен индекс 0, всем точкам справа от нее будут присвоены положительные индексы от 1 до n, где n – длина бинарной строки. После этого значение точки с индексом i, где i – нечетное число, меняется на значение точки с индексом i+1. Пример подобной мутации относительно 4-го бита исходной строки приведен в табл. 2.

Таблица 2 – Многоточечный оператор мутации

До мутации	1	0	1	0	1	0	1	0
После мутации	1	0	1	0	0	1	0	1

Метод инверсии случайных бит. Определяется количество случайных бит k, которые у данного детектора будут подлежать мутации, затем данные биты инвертируются. В таблице 3 приведен набор бит до мутации и после мутации, k = 1, мутация первого бита.

Таблица 3 – Набор бит до мутации и после мутации

До мутации	1	0	1	0	1	0	1	0
После мутации	0	0	1	0	0	1	0	1

Метод инверсии бит относительно выбранной точки разреза заключается в выборе точки разреза и инвертировании значения бит справа от выбранной точки.

3.3 Селекция

При моделировании искусственной иммунной системы оператор селекции использован с целью выбора наиболее эффективных детекторов – родителей, на основе которых будут созданы новые детекторы – потомки. Существует множество методов селекции, рассмотрим основные, смоделированнные в программном комплексе:

– метод колеса рулетки;

– элитная селекция;

– турнирная селекция.

При использовании метода рулетки вероятность того, что данный детектор попадет в итоговую выборку определяется по формуле (1)

$$P = \frac{f(i)}{\Sigma f(i)}, \quad (1)$$

где $f(i)$ – значение функции аффинности для данного детектора, $\sum f(i)$ – сумма значений функций аффинности всех детекторов.

При элитной селекции в итоговую выборку с вероятностью 1 попадают образцы с наибольшим значением функции аффинности. В турнирной селекции выбирается некоторая случайная выборка, из которой будут выбраны образцы с наибольшей функцией аффинности.

Под функцией аффинности или показателем приспособленности будем понимать отношение, представленное формулой (2):

$$K = \frac{R(i)}{\Sigma R(i)}, \quad (2)$$

где $R(i)$ – количество различных бит в соответствующих позициях детектора и патогена, $\sum R(i)$ – количество бит у патогена.

Соотнесем описанные выше компоненты иммунной системы со смоделированными компонентами и свойствами искусственной иммунной системы. Результат представим в виде таблицы 4.

Таблица 4 – Соответствие компонентов и свойств ИС и ИИС

Иммунная система	ИИС
В – лимфоцит	Детектор
Т – лимфоцит	Детектор
Тимус	Отсутствует
Рецептор	HEX–строка (возможно применение бинарных строк)
Патоген	Несобственная HEX–строка
Химическая связь между рецептором и патогенном	Расстояние Левенштейна/ Подсчет r – максимального числа смежных бит
Клонирование лимфоцитов	Создание новых детекторов на основе селекции
Обнаружение патогенна	Превышение порогового уровня у детектора
Кроссинговер	Операция над строками
Мутация	Операция над строкой
Селекция	Отношение функции аффинности детектора к сумме всех функций аффинности
Чувствительной иммунной системы	Различные весовые значения у ребер и вершин графа
Память	Долгоживущие детекторы, база данных детекторов
Уничтожение патогенов	Иммунный ответ
Созревание Т–хелперов	Период обучения, взаимодействие детекторов с собой
Поддержание численности популяции	Программное уничтожение детекторов

4 Методика обнаружения вирусов на основе ИИС и ГА

1) Генерация начальной популяции детекторов. На первом шаге алгоритма определяется длина строки детектора и происходит ее генерация случайным образом на основе алгоритма Блюм – Блюма – Шуба.

2) Иммунологическое обучение детекторов. Задается период обучения детекторов Т, во время которого детекторы взаимодействуют друг с другом и гарантированно невредоносными файлами. Детекторы, переходящие в активное состояние, подлежат «программной смерти», таким образом детекторы обучаются быть толерантными по отношения к себе.

3) Функционирование детектора в системе:

3.1 Детектор выбирает случайным образом файл, который не был им проверен, и начинает его сканирование.

3.2 Если пороговое значение превышено не было, то переход на шаг 3.1.

3.3 Если было превышено пороговое значение, то файл будет помечен как вирус, детектор начнет процесс клонирования (создания копий, у которых цикл жизни закончится, если они не попадут в выборку детекторов с максимальной функцией аффинности) и мутации себя.

3.4 Клоны начинают проверку файлового пространства до тех пор, пока все модификации вируса не будут нейтрализованы.

3.5 Выборка среди всех детекторов таких детекторов, у которых значение функции аффинности максимально, и дальнейшие занесение их в постоянную память иммунной системы.

4) Вычисление функции аффинности для каждого детектора на основе формулы.

5) Формирование выборки детекторов, к которым будет применен оператор многоточечного кроссинговера.

6) Кроссинговер. Получение новой выборки детекторов, к которой будет применен оператор мутации.

7) Мутация. Задается число k – количество случайных бит, которые будут выбраны случайным образом у детектора и будут инвертированы.

8) Поддержание постоянной численности популяции детекторов, «программная смерть» для детекторов с низким показателем функции аффинности.

9) Занесение эффективных детекторов в память.

10) Возврат на шаг 3.

Рисунок 1 – Процедура повышения меры аффинности популяции детекторов

Процедура повышения меры аффинности популяции детекторов представлена на рисунке 1. На рисунке 2 представлен жизненный цикл детектора.

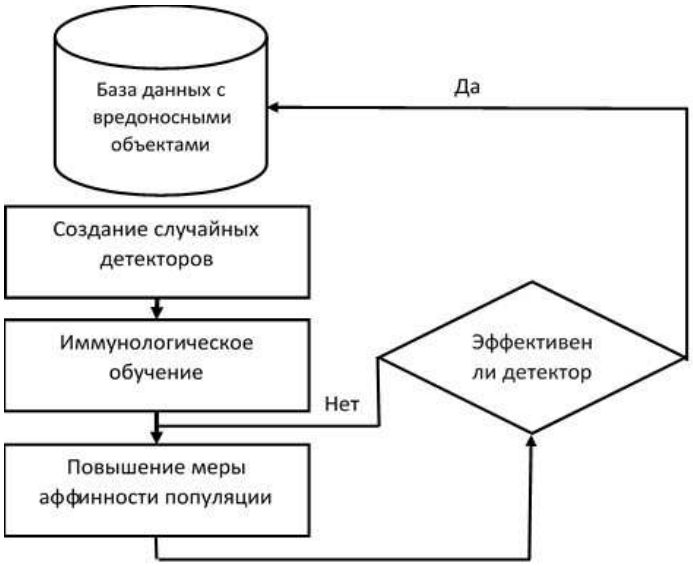

Рисунок 2 – Жизненный цикл детектора

Таким образом, описанная искусственная иммунная система представляет собой высоко распараллеленную и распределенную среду, большое количество детекторов (лимфоцитов) делают систему стабильной, а отсутствие единой точки выхода из строя (является децентрализованной) делает ее надежной. Фундаментальные компоненты и свойства ИИС говорят о ее применимости для задач распознавания полиморфных вирусов. Также стоит отметить результативность использования класса генетических алгоритмов: применение концепций генетических операторов позволяет увеличить скорость обучения детекторов, а, следовательно, и эффективность работы самой системы.

Литература

1. Частикова В.А. Идентификация механизмов реализации операторов генетического алгоритма в экспертных системах продукционного типа //Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2012. № 75. С. 308-320.

2. Частикова В.А., Картамышев Д.А. Искусственные иммунные системы: основные подходы и особенности их реализации //Научные труды Кубанского государственного технологического университета. 2016. № 8. С. 193-208.

3. Частикова В.А. Исследование основных параметров генетического алгоритма метода генетических схем в интеллектуальных системах, основанных на знаниях //

Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2011. № 69. С. 151-163.

4. Частикова В.А. Оптимизация процессов поиска решений в интеллектуальных системах обработки экспертной информации на основе генетических алгоритмов // Диссертация на соискание ученой степени кандидата технических наук / Кубанский государственный технологический университет. Краснодар, 2005.

5. de Castro Leandro N. Artificial Immune Systems: A New Computational Intelligence Approach. - Springer, 2002.

6. Kephart, J. O. (1994). "A biologically inspired immune system for computers". Proceedings of Artificial Life IV: The Fourth International Workshop on the Synthesis and Simulation of Living Systems: 130-139, MIT Press.

7. D. Dasgupta (Editor), Artificial Immune Systems and Their Applications, Springer-Verlag, Inc. Berlin, January 1999.

8. H. Bersini, F.J. Varela, Hints for adaptive problem solving gleaned from immune networks. Parallel Problem Solving from Nature, First Workshop PPSW, Dortmund, FRG, October, 1990.

9. J.D. Farmer, N. Packard and A. Perelson, (1986) "The immune system, adaptation and machine learning", Physica D, vol.2,pp. 187-204.

References

1. Chastikova V.A. Identifikacija mehanizmov realizacii operatorov geneticheskogo algoritma v jekspertnyh sistemah produkcionnogo tipa //Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2012. № 75. S. 308-320.

2. Chastikova V.A., Kartamyshev D.A. Iskusstvennye immunnye sistemy: osnovnye podhody i osobennosti ih realizacii //Nauchnye trudy Kubanskogo gosudarstvennogo tehnologicheskogo universiteta. 2016. № 8. S. 193-208.

3. Chastikova V.A. Issledovanie osnovnyh parametrov geneticheskogo algoritma metoda geneticheskih shem v intellektual'nyh sistemah, osnovannyh na znanijah // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2011. № 69. S. 151-163.

4. Chastikova V.A. Optimizacija processov poiska reshenij v intellektual'nyh sistemah obrabotki jekspertnoj informacii na osnove geneticheskih algoritmov // Dissertacija na soiskanie uchenoj stepeni kandidata tehnicheskih nauk / Kubanskij gosudarstvennyj tehnologicheskij universitet. Krasnodar, 2005.

5 de Castro Leandro N. Artificial Immune Systems: A New Computational Intelligence Approach. - Springer, 2002.

6 Kephart, J. O. (1994). "A biologically inspired immune system for computers". Proceedings of Artificial Life IV: The Fourth International Workshop on the Synthesis and Simulation of Living Systems, MIT Press.

7 D. Dasgupta (Editor), Artificial Immune Systems and Their Applications, Springer-Verlag, Inc. Berlin, January 1999.

8 H. Bersini, F.J. Varela, Hints for adaptive problem solving gleaned from immune networks. Parallel Problem Solving from Nature, First Workshop PPSW, Dortmund, FRG, October, 1990.

9 J.D. Farmer, N. Packard and A. Perelson, (1986) "The immune system, adaptation and machine learning", Physica D, vol.2,pp. 187-204.