ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ЭЛЕМЕНТОВ ДВУКЛАССИФИКАЦИОННОЙ ИСКУССТВЕННОЙ ИММУННОЙ СИСТЕМЫ ДЛЯ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ

М.Е. Бурлаков, М.Н. Осипов

Самарский государственный аэрокосмический университет имени академика С.П. Королёва (национальный исследовательский университет) (СГАУ), Самара, Россия

В данной статье предлагается к рассмотрению двуклассификационная искусственная иммунная система. Описываются понятия базового элемента и элемента с памятью. Задается метрика между элементами системы, называемая аффинностью, с соответствующим пределом меры (порогом аффинности). Дается определение операциям клонирования и мутации. Исследуется зависимость базовых элементов и элементов с памятью при наличии метрики между ними и воздействии на них операций клонирования и мутации.

Ключевые слова: искусственные иммунные системы, мутация, клонирование, классификация электронных сообщений, обнаружение угроз, информационная безопасность, защита информации.

Введение

В современных системах по передаче информации крайне актуально стоит задача, связанная с классификацией блоков данных и электронных сообщений, передающихся от отправителю к адресату через разного рода системы (mail, web, irq и т.д.). Для ее решения применяется множество как неадаптивных (методы графов сценариев атак, методы анализа систем состояний, экспертные системы, методы на спецификациях, сигнатурные методы и т.д.), так и адаптивных (искусственные нейронные сети, искусственные иммунные алгоритмы, генетические алгоритмы и т.д.) методов [1-15].

Задача классификации электронных сообщений в рамках информационной системы сводится к отнесению входящего потока данных к соответствующему классу (например, по релевантности, по отправителю, по содержимому и объему).

Наиболее актуальным направлением в анализе и классификации сообщений выделяют классификацию по их содержимому, когда можно без знания об источнике сообщения, с определенной долей вероятности определить к какому классу она относится и либо отправить сообщение дальше, либо остановить его движение по информационной системе.

Одной из основных задач классификации блоков данных и электронных сообщений заключается в их распределении по соответствующим классам достоверности [16-19]:

1. Класс достоверной (актуальной, легитимной и т.д.) информации;
2. Класс недостоверной (неактуальной, нелегитимной и т.д.) информации.

Под **достоверной (легитимной) информацией** понимается набор данных, который не представляет из себя угрозы для информационной системы, в которой происходит ее

циркуляция, с точки зрения доступности, целостности и конфиденциальности. В противном случае информация называется недостоверной (нелегитимной).

Классический пример подобной классификации представляет из себя антиспам систему или программно-аппаратный комплекс антивирусной защиты, задача которых также сводится к определению достоверности входящего в информационную систему сообщения по его содержимому. Как было обозначено выше, существует большое количество как адаптивных, так и неадаптивных алгоритмов, способных классифицировать блоки данных (электронные сообщения) по содержимому. В работе [20] предлагается для решения данного рода задачи использовать двуклассификационную искусственную иммунную систему.

Двуклассификационная искусственная иммунная система

Двуклассификационная искусственная иммунная система (2КИИС) - адаптивный алгоритм с учителем, позволяющий классифицировать блоки данных (электронные сообщения) на два класса: класс д*остоверных* сообщений и класс *недостоверных* сообщений.

Алгоритм двуклассификационной искусственной иммунной системы разрабатывался как аналог биологической иммунной системы. Базовыми понятиями 2КИИС являются: β-элемент (базовый элемент) и β^m-элемент (элемент с памятью) – аналоги *B*-лимфоцита и *B*-лимфоцита с памятью в биологических системах.

B-лимфоцит и *B*-лимфоцит с памятью, с биологической точки зрения, являются «флагами безопасности», в случае изменения которых подается сигнал организму о необходимости создания антител, способных бороться с возникшей угрозой (антигенами). Перевод *B*-лимфоцита в *B*-лимфоцит с памятью обеспечивается за счет иммунной системы в случае, если последний успешно осуществляет борьбу (с большей вероятностью реагирует) с той или иной угрозой. Перевод *B*-лимфоцита в *B*-лимфоцит с памятью производится иммунной системой за счет наличия некоторого весового параметра. Принцип следующий, чем эффективнее лимфоцит действует, тем выше его «вес», и наоборот. В случае, если лимфоцит не обеспечивает эффективный уровень обнаружения угрозы, иммунная система понижает значение весового параметра, вплоть до достижения момента, когда он может быть устранен из системы вообще.

Исходя из представленной аналогии *B*-лимфоцит соответствует β-элементу, а *B*-лимфоцит с памятью - β^m-элементу.

Также, по аналогии, неотъемлемой частью β-элемента является понятие *возраста (мощности, веса)* – и как ненулевого значения, характеризующего вес элемента и выполняющего условие "гибели" элемента в случае его нулевого значения, и как оценки его эффективности. В 2КИИС также вводятся операции **мутации** и **клонирования**.

Под **мутацией элемента** понимается процесс случайного изменения части его структуры. Мощность элемента конечна, и процесс мутации затрагивает только определенное количество сущностей. Данное определенное количество задается в процентном соотно-

шении на этапе инициализации алгоритма 2КИИС. Операция мутации (*Mutating*) эквивалентна записи:

$$Mutating(\beta) = \sum \beta_i \bigcap \alpha_j, \alpha = \sum \alpha_j, \\ |\alpha| >> |\beta|, i = \lfloor m \cdot j \rfloor, m \in [0,1] \quad (1)$$

где знак пересечения эквивалентен операции замены,

α – множество сущностей из β – элемента,

m – коэффициент мутации.

Операция клонирования (*Cloning(*β*)*) в рамках искусственной иммунной системы необходима для создания новых β-элементов. Данная операция применяется сразу после операции мутации и заключается в копировании новых полученных β-элементов во множество уже имеющихся элементов. Важным параметром операции клонирования является коэффициент клонирования, который говорит, сколько элементов относительно исходного будет создано.

Таким образом, мутация обеспечивает создание новых элементов, потенциально способных распознавать новые угрозы, а клонирование отвечает за их «распространение» в рамках системы.

Именно благодаря этим операциям и обеспечивается вариативность между β^m-элементами и простым β-элементами. Вариативность заключается в процессе формирования: β^m-элемент - элемент с памятью, это наиболее устоявшийся "слепок" угрозы, который эффективно способен ее детектировать. Если простой β-элемент получается путем комбинации случайного выбора слов из электронного сообщения и применения операций клонирования и мутации, то β^m-элемент это элемент, полученный из итеративного измерения эффективности β-элемента при применении на выборочном множестве электронных сообщений.

Для расчета расстояния между β-элементами введем понятие метрики, называемое **аффинностью элементов**.

Аффинность двух элементов в рамках 2КИИС задается как отношение между количеством общих сущностей, из которых состоят эти элементы к норме. Норма двух элементов – минимальное количество сущностей, образующих каждый из элементов. Следовательно, аффинность (*Affinity* или α) может быть записана в виде следующего соотношения:

$$Affinity(\beta^1, \beta^2) = \alpha(\beta^1, \beta^2) = \\ = \frac{Count(\sum \beta_i^1 \bigcap \sum \beta_j^2)}{\min(|\beta_i^1|, |\beta_j^2|)}, i \neq j \quad (2)$$

где *i, j* – порядок следования сущностей в элементе;

Count – функция количества пересечений элементов;

$|\beta_j|$ - мощность элемента (количество сущностей в элементе).

Если значение аффинности между антигеном (угрозой, потенциальное сообщение из класса недостоверной информации) и любым β-элементом выше некоторого, заранее заданного на этапе инициализации, порогового значения (**порог аффинности**), то это означает, что β-элемент распознал угрозу, и далее система маркирует этот блок (электронное сообщение) как недостоверный, в противном случае блок маркируется как достоверный.

Наличие метрики в виде аффинности позволяет обнаружить угрозу и тем самым осуществить процесс первичной классификации. С другой стороны, стоит вопрос связанный с исследованием зависимости между процессами формирования β-элементов и β^m-элементов. Другими словами, существует задача, которая требует однозначно сказать – существует ли зависимость между базовыми (простыми) β-элементами и β^m-элементами с памятью при наличии метрики в виде аффинности и существовании операций клонирования и мутации в рамках решения задачи классификации электронных сообщений на два класса: класс достоверных сообщений и класс недостоверных сообщений.

Для решения поставленной задачи о нахождении зависимости между β-элементами и β^m-элементами докажем Утверждение 1.

Утверждение 1.

Для любого $\varepsilon > 0$ существует $\lambda \in [0,1]$ такое, что $|\beta| = \lambda|\beta^m|$.

Доказательство.

Введем ряд обозначений.

Рассмотрим исходное множество сообщений (электронных сообщений, блоков данных) S из которых происходит инициализация β – элементов, где $|S| = s$, $s>0$ и $s \in N$. Пусть k – возраст β – элемента, и n – порог, при котором $\beta \to \beta^m$ (порог аффинности), а $r = n - k$, $r>0$, их разница. В момент инициализации в системе находится ровно $\bigcup_{i=1}^{s} \beta(k)$ – элементов (слов текста электронного сообщения).

Пусть α – значение аффинного расстояния функции *Affinity()* двух элементов. P_α – вероятность появления двух таких β-элементов, аффинное расстояние между которыми будет больше или равно α, а p_i – ее соответствующее значение. За l обозначим коэффициент клонирования при применении функции клонирования *Cloning(*β*)* (далее кратко обозначим как функцию *C(*β*)*). За переменную m – обозначим коэффициент мутации с соответствующими вероятностными характеристиками возникновения новых β-элементов при применении операции мутации *Mutating(*β*)* (далее кратко обозначим как функцию *M(*β*)*).

Общий алгоритм получения β и β^m-элементов выглядит следующим образом:

Шаг 1. Инициализация множества β-элементов из множества S;

Шаг 2. Применение функции мутации и клонирования к множеству β-элементов;

Шаг 3. Расчет вероятности P_α к полученному множеству β-элементов;

Шаг 4. Переход к Шагу 2.

Рассмотрим процесс итеративно. На Шагах 1 и 2 к последовательности β-элементов применяются операции клонирования и мутации:

$$C(M(\bigcup_{i=1}^{s}\beta(k))) \tag{3}$$

На Шаге 3, к полученному множеству β-элементов применяется расчет вероятности P_α. Таким образом, будет получено новое множество β-элементов, возрасты которых могут быть не равны между собой вследствие наличия порога аффинности. Далее проделываем ту же последовательность начиная со второго шага. Алгоритмически, последовательность операций складывается из конечного набора итераций.

Итерация 1.

$$\begin{aligned} &M_1 = C(M(\bigcup_{i=1}^{s}\beta(k))) \mid P_\alpha = \\ &= \bigcup_{i_1=1}^{b_{11}\le s_1}\beta_{i_1}(k-1)\bigcup_{i_2=1}^{b_{12}\le s_1-b_{11}}\beta_{i_2}(k+1) \end{aligned} \tag{4}$$

На первой итерации было получено множество M_1, состоящее из двух подмножеств β-элементов: возраст которых равен k-1 (если аффинное расстояние меньше α) и k+1 (если аффинное расстояние больше или равно α), и где $s_1 \geq s$ из-за операции клонирования. Мощность множества M_1 конечна и равна (вследствие конечности множества S):

$$M_1 = p_1 l_1 m_1 \mid \beta(k-1) \parallel \beta(k+1) \mid \tag{5}$$

где p_1 – значение вероятности P_α ,

l_1 – коэффициент клонирования на первом шаге,

m_1 – коэффициент мутации на первом шаге,

$\beta(k-1)$ – количество β-элементов, возраст которых равен k-1,

$\beta(k+1)$ – количество β-элементов, возраст которых равен k+1.

Итерация 2 рассчитывается исходя из повторного применения операций клонирования и мутации с учетом результатов полученных на предыдущей итерации, т.е.:

$$\begin{aligned} &M_i = C(M(M_{i-1})) \mid P_\alpha = \\ &C(M(C(M(M_{i-2})))) \mid P_\alpha = C(M\ldots(M_1)\ldots) \mid P_\alpha \end{aligned} \tag{6}$$

Таким образом, результат итерации 2 имеет вид:

$$M_2 = C(M(M_1)) \mid P_\alpha = \bigcup_{i_1=1}^{b_{21} \le b_{11}} \beta_{i_1}(k-2) \times \\ \times \bigcup_{i_2=1}^{b_{22} \le b_{11} - b_{21}} \beta_{i_2}(k) \bigcup_{i_3=1}^{b_{23} \le b_{12}} \beta_{i_3}(k) \bigcup_{i_4=1}^{b_{24} \le b_{12} - b_{23}} \beta_{i_4}(k+1) \quad (7)$$

а мощность множества M_2 конечна и равна:

$$| M_2 | = p_2 l_2 m_2 \mid \beta(k-2) \parallel \beta(k) \parallel \beta(k+1) \mid \quad (8)$$

где параметры по аналогии с Итерацией 1. Для r-ой итерации, множество M_r примет вид:

$$M_2 = L(M(M_{r-1})) \mid P_\alpha = \bigcup_{i_1=1}^{b_{r1} \le b_{(r-1)1}} \beta_{i_1}(k-r) \times \\ \times \bigcup_{i_2=1}^{b_{r2} \le b_{(r-1)1} - b_{r1}} \beta_{i_2}(k-r+1) \ldots \bigcup_{i_2 r=1}^{b_{r2} r \le b_{(r-1)1} - \sum_{j=1}^{r-1} b_{rj}} \beta_{i_2 r}(k+r) \quad (9)$$

Именно на данном этапе происходит получение β^m-элементов вследствие наличия β-элементов возраст которых равен $n = k + r$ – порога, при котором $\beta \to \beta^m$.

Мощность множества M_r конечна и равна:

$$| M_r | = p_r l_r m_r \mid \beta(k-r) \parallel \beta(k-r+1) \mid \ldots \\ \mid \beta(k+r-1) \parallel \beta(k+r) \mid = p_r l_r m_r \mid \beta \parallel \beta^m \mid \quad (10)$$

где |β| - мощность всех β-элементов:

$|\beta^m|$ - мощность все β^m-элементов.

Зависимость между количеством β и β^m-элементов есть величина (коэффициент), зависящая от коэффициентов клонирования, мутации и аффинного расстояния больше или равной заданной величины. Таким образом, утверждение, что для любого ε > 0 существует $\lambda \in [0,1]$ такое, что $|\beta| = \lambda|\beta^m|$ справедливо.

Вторым важным вопросом является установление зависимости количества простых β-элементов и β^m-элементов с памятью от тех множеств, в рамках которых происходит их формирование. То есть, стоит задача определения зависимости мощности β-элементов и β^m-элементов двух конечных множеств электронных сообщений, имеющих отношение вложенности между собой. Для этого сформулируем Утверждение 2.

Утверждение 2.

Для любых двух конечных множеств S_1 и S_2, где $S_1 \subseteq S_2, | S_1 | \le | S_2 |$, справедливо $|\beta_1| \le |\beta_2|$ *и* $|\beta^m{}_1| \le |\beta^m{}_2|$.

Доказательство.

Доказательство сводится к повторению итераций из Утверждения 1, но уже для двух множеств S_1 и S_2 со сравнением полученных результатов.

Таким образом, наличие прямой зависимости простых β-элементов от β^m-элементов позволяет оценить количество порождаемых β-элементов от выбранного множества и в дальнейшем прогнозировать количество β^m-элементов с целью проектирования информационных систем классифицирующих электронные сообщения лимитированного масштаба.

Заключение

Таким образом, существует прямая связь между β-элементами и βm-элементами в рамках конечного множества электронных сообщений, что позволяет более точно спрогнозировать количество выделяемой памяти для создания и хранения β-элементов и βm-элементов в рамках двуклассификационной искусственной иммунной системы. Построенные на этом основании информационные системы, способны классифицировать электронные сообщения на достоверные и недостоверные классы, обеспечивая тем самым процесс непрерывного обнаружения вторжений на основе их содержимого.

Литература

1. Васильев, В.И. Интеллектуальные системы защиты информации / В.И. Васильев - М.: Машиностроение, - 2012. – С. 20-22.
2. Vacca, J.R. Computer and Information Security Handbook / John R. Vacca // Newnes. – 2012. – P. 330-335.
3. Nunes, L. Artificial Immune Systems: A New Computational Intelligence Approach / Leandro Nunes de Castro, Jonathan Timmis // Springer Science & Business Media. – 2002. – P. 2-4.
4. Хайкин, С. Нейронные сети: полный курс, 2-е издание / С. Хайкин - М.: Издательский дом Вильямс, 2008. – С. 32-34.
5. Abe S. Support Vector Machines for Pattern Classification / Shigeo Abe // Springer Science & Business Media. – 2005. – P. 39-40.
6. Kollias, S. Artificial Neural Networks / Stefanos Kollias // Springer Science & Business Media. – 2006. – P. 161-162.
7. Искусственные иммунные системы и их применение // под ред. Д. Дасгупты: пер. с англ. – М.: ФИЗМАТЛИТ, 2006. – С. 340-344.
8. Tarakanov, A.O. Immunocomputing: principles and applications / A.O. Tarakanov, V.A. Skormin, S.P. Sokolova, // Springer Verlag, New York, 2003.
9. Vacca, J.R. Computer and Information Security Handbook / John R. Vacca // Newnes. – 2012. – P. 330-335.
10. Borger, E. The Abstract State Machines Method for High-Level System Design and Analysis / Egon Borger // Dipartimento di Informatica, Universita di Pisa. – 2007. – P. 30-35.
11. Shim, J.K. Information Systems and Technology for the Noninformation Systems Executive / Jae K. Shim // CRC Press. – 2000. – P. 230-235.
12. Lunt, T.F. A real-time intrusion-detection expert system (IDES) / Teresa F. Lunt, Ann Tamaru, Fred Gilham // Final Technical Report. – 1992. – P. 10-13.
13. Бурлаков, М.Е. Метод фильтрации входящего трафика на основе двухслойной рекуррентной нейронной сети // Ползуновский вестник №3/2. - Алтайский государственный технический университет им. И.И. Ползунова, 2012. – С. 215-219.
14. Бурлаков М.Е., Осипов М.Н. Аудит безопасности локальной вычислительной сети с помощью динамической системы на нейронах с реакцией на последовательности. // Информационное противодействие угрозам терроризма. 2013. № 20. С. 166-170.

15. Бурлаков М.Е. Обзор базовых алгоритмов искусственных иммунных систем на клонально-селективной теории // Сборник статей Международной научно-практической конференции "Приоритетные направления развития науки". - УФА Аэтерна, 2014. - 185-192 с.
16. Бурлаков М.Е. О некоторых моделях оптимизации искусственной нейронной сети генетическими алгоритмами // Перспективные информационные технологии (ПИТ-2014): труды Международной научно-технической конференции. - Самара: Издательство Самарского научного центра РАН, 2014. - 99-105 с.
17. Delvin, D. Satisfiability as a Classification Problem // University College Cork. URL: http://www.cs.ucc.ie/~osullb/pubs/classification.pdf (дата обращения 03.01.2016)
18. Fernandez-Delgado, M. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems // University of Santiago de Compostela. URL: http://jmlr.csail.mit.edu/papers/volume15/delgado14a/delgado14a.pdf.
19. Schapire, R. Machine Learning Algorithms for Classification // Princeton University. URL: http://www.cs.princeton. edu/~schapire/talks/picasso-minicourse.pdf.
20. Бурлаков, М.Е. Двухклассификационная искусственная иммунная система // Вестник Самарского государственного университета №7(118), 2014 - С. 207-221.