УДК 621.391.16

НЕЙРОСЕТЕВОЙ АЛГОРИТМ ОБРАБОТКИ ИНФОРМАЦИИ ДЛЯ ПРОГНОЗИРОВАНИЯ НАДЕЖНОСТИ ПРОГРАММНОГО ОБЕСПЕЧЕНИИЯ

А. С. Вялых, С. А. Вялых, А. А. Сирота

Воронежский государственный университет

Поступила в редакцию 24.06.2013 г.

Аннотация. В статье описывается нейросетевой алгоритм обработки информации, позволяющий прогнозировать динамику обнаружения уязвимостей в программном обеспечении, которые влияют на надежность работы информационных систем.
Ключевые слова: нейронная сеть, надежность, уязвимость, программное обеспечение, информационная система.
Annotation. In the article we describe neuronetwork algorithm of the information processing, allowing to predict detection dynamics of vulnerabilities in the software, which influence reliability of information systems work.
Keywords: neural network, reliability, vulnerability, software, information system.

ВВЕДЕНИЕ. Одним из основных факторов, влияющих на надежность информационных систем (ИС) и информационных технологий в условиях целенаправленных негативных воздействий, является наличие уязвимостей в программном обеспечении (ПО), установленном в ИС. При этом для анализа будущего состояния ИС необходим прогноз в отношении динамики обнаружения уязвимостей, которые можно использовать для нарушения целостности и доступности информации в ИС. На данный момент существует ряд аналитических прогностических моделей обнаружения уязвимостей [1]. Хотя исследования данных моделей и показали, что в большинстве случаев наилучшими прогностическими способностями обладает логистическая модель Алхазми-Малаийа [1], те же исследования показали, что в ряде случаев более эффективными оказываются другие аналитические модели. В [2] было показано, что число обнаруженных в ПО за месяц уязвимостей зависит не только от времени существования ПО, но и от того, какой это конкретный месяц года, что аналитические модели не учитывают. И более того, динамика обнаружения уязвимостей имеет случайную составляющую, что также не учитывается аналитическими моделями. В итоге данные модели способны прогнозировать только усредненные тенденции в изменении скорости обнаружения уязвимостей, когда по факту их значительно больше [3]. Ошибка даже всего лишь на одну уязвимость может привести к заведомо недостоверной оценке надёжности программного обеспечения. В связи с этим целью данной работы является повышение точности прогнозирования надежности программного обеспечения на основе использования нейросетевых алгоритмов, которые позволяют учесть вышеназванные факторы.

ОПИСАНИЕ АЛГОРИТМА. Для прогнозирования обнаружения уязвимостей при помощи искусственных нейронных сетей [4,5] использовались следующие исходные данные. Прогноз производился для операционной системы Windows XP на 1 и 6 месяцев вперед. Данные сроки прогноза актуальны как для разработчиков, так и для пользователей ПО, так как позволяют в краткосрочной и среднесрочной перспективе первым грамотно распределить ресурсы между разработкой нового ПО и сопровождением различного старого ПО, а вторым идентифицировать слабые места в ИС (ПО, в котором будет найдено большее число уязвимостей), позволяющие нарушить ее работу, и принять соответствующие меры. Прогноз осуществлялся для случаев, когда известны данные за 96, 102, 108, 114, 120 и 126 меся-

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 13-01-97507 р_центр_ а

цев. Для прогноза использовались данные по уязвимостям для Windows XP из National Vulnerability Database [3]. Анализ данных проводился в среде Matlab с использованием компонентов входящей в нее подсистемы Neural Network Toolbox. Обработка осуществлялась в два этапа.

На первом этапе осуществлялась предварительная обработка данных, полученных в моменты времени $t^{(1)},...,t^{(P)}$, обеспечивающая их сглаживание и интерполяцию в виде непрерывной функциональной зависимости от времени. При проведении предварительной обработки осуществлялось восстановление зависимости на основе взвешенной суммы радиально-базисных функций

$$F(t) = \sum_{i=1}^{K} w_i \varphi_i(t) = w^T \varphi(t),$$

$$\varphi_i(t) = \varphi\left(\left\|t - u_i\right\|\right) =$$

$$= \exp\left[-\frac{\left(\left\|t - u_i\right\|\right)^2}{2\sigma_i^2}\right],$$

где $\varphi_i(t)$ – i-я радиально-базисная функция; u_i – центр i-ой радиально-базисной функции; σ_i – параметр влияния i-ой радиально-базисной функции; w_i – соответствующий весовой коэффициент этой функции; K – количество используемых функций.

При вычислении коэффициентов ряда проводилось решение переопределенной системы линейных уравнений на основе метода регуляризации А.Н. Тихонова с априорным решением, в качестве которого использовалась логистическая модель Алхазми-Малаийа [1], на основе следующих соотношений:

$$Gw = d,$$

$$w = w^{(\alpha)} + \left(G^T G + \alpha I\right)^{-1} G^T (d - Gw^{(\alpha)}),$$

$$G = \left\|g_{p,i}\right\|,$$

$$g_{p,i} = \left\|\varphi_i\left(t^{(p)}\right)\right\|,$$

$$p = \overline{1,P},$$

$$i = \overline{1,K},$$

$$K < P,$$

где G – матрица Грина [4], являющаяся в данном случае прямоугольной; $d = (d^{(1)},...,d^{(P)})^T$ – целевой вектор, определяемый из исходного множества аппроксимируемых данных; $w^{(\alpha)}$ – априорное решение; α – параметр регуляризации, выбираемый одним из стандартных методов; I – единичная матрица размера $K \times K$; P – число моментов времени, для которых рассчитываются значения базисно радиальных функций (в рассматриваемом случае – число месяцев, по которым есть данные по количеству новых обнаруженных уязвимостей).

На рисунке 1 показаны результаты предварительной обработки, реализованной в среде Matlab на основе представления функциональной зависимости в виде взвешенной суммы $K = 20$ радиально-базисных функций. Весовые коэффициенты разложения получены методом регуляризации по Тихонову с параметром регуляризации $\alpha = 0,01$.

На втором этапе обработки осуществлялось прогнозирование сглаженных и интерполированных данных с использованием комитета из 10 искусственных двухслойных нейронных сетей прямого распространения с сигмоидной функцией активации в виде гиперболического тангенса для 1-го слоя (в Matlab функция tansig) и линейной функций активации для 2-го слоя (в Matlab функция purelin). Для построения прогнозирующего алгоритма проводилось обучение каждой нейронной сети при помощи функции, которая модифицирует веса и смещения в соответствии с методом шкалированных связанных градиентов (в Matlab функция trainscg), обеспечивающее восстановление нелинейной авторегрессионной зависимости разницы между очередным (прогнозируемым) значением и предыдущим значением анализи-

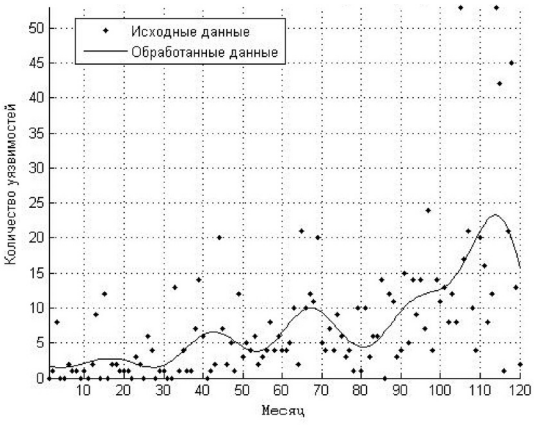

Рис. 1. Результаты сглаживания и интерполяции данных по уязвимостям (влияющим на надежность работы ИС), обнаруженным в Windows XP за 120 месяцев

руемого процесса от Nu предшествующих значений. В качестве итогового результата прогноза за каждый месяц бралось среднее значение между прогнозами 10 нейронных сетей.

На рисунке 2 приведены результаты прогноза для Windows XP по данным, полученным в ходе предварительной обработки, на период с 121 по 126 месяц для значений параметра $Nu = 40$.

Предложенный способ прогнозирования обнаружения уязвимостей сравнивался с прогнозом, получаемым при помощи логистической модели Алхазми-Малаийа [1]. Для этого вычислялось среднее абсолютное отклонение прогноза от реальных данных. Результаты сравнения приведены в таблице 1.

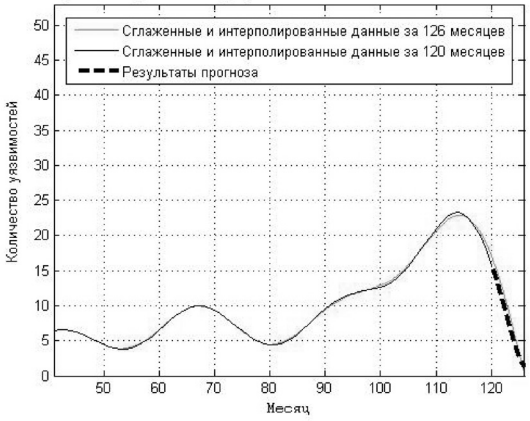

Рис. 2. Результаты прогноза обнаружения уязвимостей (влияющих на надежность работы ИС) в Windows XP на 6 месяцев при известных данных за 120 месяцев

Таблица 1

Среднее абсолютное отклонение одномесячного и полугодового прогноза обнаружения уязвимостей в Windows XP от реальных данных

Известные данные (время жизни ПО), месяцы	Среднее абсолютное отклонение прогноза от реальных данных, уязвимости			
	Логистическая модель Алхазми-Малаийа		Нейронная сеть (априорное решение – логистическая модель Алхазми-Малаийа)	
	Прогноз на 1 месяц	Прогноз на 6 месяцев	Прогноз на 1 месяц	Прогноз на 6 месяцев
126	11,15	11,99	3,67	12,76
120	9,82	7,98	8,97	3,62
114	28,79	15,33	27,53	15,27
108	8,85	10,82	9,62	10,83
102	0,31	10,2	0,31	10,2
96	14,8	6,77	14,53	5,26
Среднее значение	12,29	10,52	10,77	9,66

ЗАКЛЮЧЕНИЕ. Результаты сравнения показывают, что прогноз динамики обнаружения уязвимостей в операционной системе Windows XP, влияющих на надежность работы ИС в условиях целенаправленных негативных воздействий, при помощи нейронной сети с использованием в качестве априорного решения логистической модели Алхазми-Малаийа точнее, чем прогноз при помощи этой модели на 8% при периоде прогноза в 6 месяцев и на 12 % при периоде прогноза в 1 месяц. При этом наибольший выигрыш нейросетевой алгоритм (почти в 3 раза) показывает при максимальном количестве обучающих данных (полученных за 126 месяцев жизненного цикла программного обеспечения) и периоде прогноза в 1 месяц. Разница в 8–12 % между нейросетевым алгоритмом и аналитической моделью прогнозирования обнаружения уязвимостей при средней скорости открытия уязвимостей (приблизительно 8 уязвимостей в месяц для Windows XP) в среднем означает неточность в 1 уязвимость в месяц, что в ряде случаев недопустимо (особенно с учетом того, что срок закрытия уязвимости может составлять месяц и более), так как наличие даже одной уязвимости в ИС позволяет нарушить ее работу, а каждая новая уязвимость в ИС предоставляет дополнительные возможности для этого.

СПИСОК ЛИТЕРАТУРЫ

1. *Alhazmi O.H.* Modeling the Vulnerability Discovery Process / O.H. Alhazmi, Y.K. Malaiya. – Proc. Int. Symp. Software Reliability Eng, Nov. 2005, pp. 129–138.

2. *Joh H.* Seasonal Variation in the Vulnerability Discovery Process / H. Joh, Y.K. Malaiya. – Proc. 2nd IEEE Int. Conf. Software Testing, Verification, and Validation, April 2009, pp. 191–200.

3. National Vulnerability Database. – http://nvd.nist.gov, 2013.

4. *Осовский С.* Нейронные сети для обработки информации / С. Оссовский. – М.: Финансы и статистика, 2002. –344 с.

5. *Хайкин С.* Нейронные сети: полный курс, 2-е изд., испр. : Пер. с англ. / С. Хайкин. — М.: ООО "И.Д. Вильямс", 2006. — 1104 с.

Вялых Александр Сергеевич – аспирант кафедры технологий обработки и защиты информации ФКН ВГУ, Воронежский государственный университет. E-mail: alexandervyalih@gmail.com

Вялых Сергей Ариевич – кандидат технических наук, доцент кафедры технологий обработки и защиты информации ФКН ВГУ, Воронежский государственный университет. E-mail: vyalyh@govvrn.ru

Сирота Александр Анатольевич – доктор технических наук, профессор кафедры технологий обработки и защиты информации ФКН ВГУ, Воронежский государственный университет. E-mail: sir@cs.vsu.ru

Vyalykh A. S.– post-graduate student, chair of information processing and security technologies, Voronezh State University. E-mail: alexandervyalih@gmail.com

Vyalykh S. A. – Cand.Tech.Sci., senior lecturer, chair of information processing and security technologies, Voronezh State University. E-mail: vyalyh@govvrn.ru

Sirota A. A. – Doct.Tech.Sci., professor, chair of information processing and security technologies, Voronezh State University. E-mail: sir@cs.vsu.ru