ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА КОМП'ЮТЕРНА ТЕХНІКА

УДК 004.492.3

О. С. Савенко, канд. техн. наук, доц.**;**
С. М. Лисенко

ПОБУДОВА АДАПТИВНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ДІАГНОСТУВАННЯ КОМП'ЮТЕРНИХ СИСТЕМ НА НАЯВНІСТЬ ТРОЯНСЬКИХ ПРОГРАМ

Розроблено адаптивну інформаційну технологію діагностування комп'ютерних систем на наявність троянських програм, суть якої полягає у використанні поведінкових моделей класів троянських програм, відмінністю її від відомих є те, що процес діагностування не потребує побудови баз сигнатур, дає змогу виявляти нові невідомі троянські програми та підвищує достовірність і ефективність процесу діагностування.

Вступ

Об'єднання комп'ютерних систем (КС) в локальні мережі та їх підключення до глобальної мережі Internet створює проблеми, пов'язані з їх функціонуванням та використанням. Серед них вагоме місце займає шкідливе програмне забезпечення (ШПЗ), яке призводить до неправильного функціонування програмного та апаратного забезпечення. Аналіз ситуації щодо ШПЗ показує інтенсивне зростання чисельності троянських програм (ТП), здатних виконувати в КС деструктивні або шкідливі дії. Розробники ТП знаходять нові способи їх потрапляння в КС, застосовують маскування від антивірусного ПЗ, вдосконалюють код ТП. В свою чергу, розробники антивірусного ПЗ постійно вдосконалюють інформаційні технології діагностування КС, оновлюють антивірусні бази, застосовують сучасні механізми виявлення ШПЗ. Проте, наявні факти викрадення конфіденційної інформації та здійснення деструктивних дій в КС, в яких встановлене антивірусне ПЗ, свідчать про недоліки відомих технологій діагностування КС на наявність ТП, які орієнтовані на виявлення відомих ТП але не повністю адаптовані до розпізнавання нових ТП.

Аналіз найпоширеніших інформаційних технологій (ІТ) антивірусного діагностування (АД) КС, якими є сигнатурний аналіз і метод контрольних сум, не здатні виявляти нові ТП, що суттєво знижує достовірність та ефективність діагностування. ІТ, які ґрунтуються на евристичному аналізі, мають високу ймовірність хибних спрацювань [1].

Мета роботи та постановка задачі

Новітніми підходами до діагностування КС на наявність невідомих ТП є залучення компонентів штучного інтелекту (ШІ), які дозволяють спростити обчислювальну складність процесу діагностування та надати властивості адаптивності ІТ, яка здійснює АД.

Таким чином, постає задача розроблення нової ІТ, яка б давала змогу діагностувати КС на наявність ТП в режимах монітора та сканера. Антивірусний монітор повинен аналізувати поведінку програмних об'єктів і робити висновок про небезпеку інфікування КС троянськими програмами. З цією метою доцільним є залучення апарату нечіткої логіки. Антивірусний сканер має дозволити здійснювати сканування КС на предмет підміни системних файлів та інших файлів троянськими версіями і здійснити перевірку цілісності даних КС. Для цього доцільно використати алгоритми штучних імунних систем, що надасть адаптивної характеристики процесу діагностування комп'ютерних систем на наявність ТП. Розробка адаптивної інформаційної технології повинна підвищити достовірність та ефективність діагностування КС на наявність троянських програм.

Поведінкова модель троянських програм

Життєвий цикл (ЖЦ) ТП складається з етапів потрапляння на віддалену КС, активізації та виконання закладених деструктивних дій ТП, на основі чого розроблено їх поведінкову модель [2]

$$M_v = \langle \Theta, S, V, L, Aff, \varepsilon, Z \rangle, \quad (1)$$

де Θ — множина усіх троянських програм; S — етапи ЖЦ троянської програми $s_i \in S$, $i = \overline{1,3}$; $V = |V_{mp}|$ — матриця відношень m дій ТП та p портів мережних протоколів; $L = |L_{ab}|$ — матриця відношень дій $a \in A$ ТП і $b \in B$ структурних одиниць операційної системи; Aff — функція, яка визначає взаємодію між об'єктами КС і ТП ν_j, тоді множина $a \in Aff(b_i, \nu_j)$ є набором можливих дій, які ТП ν_j завдає об'єкту (об'єктам) b_i; ε — відношення між ТП та її станами, тоді для $\nu \in \Theta$ та $s \in S$, відношення $\nu \; \varepsilon \; s$ означає, що ТП ν перебуває в стані s; відношення $\nu \; \bar{\varepsilon} \; s$ означає, що ТП ν не перебуває в стані s; Z — характеристичні параметри відношень, $Z = \{z_k\}$ — вектор деструктивних дій об'єкта з нормованими пріоритетними вагами $P = \{p_k\}$ $\left(\sum p_k = 1\right)$, що враховують рівень їхньої небезпеки для КС [2]. Також в поведінкову модель ТП введено позначення $\longrightarrow$, яке задає відношення між трьома поняттями, а саме: якщо $s_i \xrightarrow{a} s_{i+1}$, то дія $a \in A$ спричиняє перехід із стану s_i в стан s_{i+1}. Тоді ТП, життєвий цикл якої має усі етапи, проходить можливий шлях: $s_0 \xrightarrow{V,L} s_1 \xrightarrow{V,L} s_2 \xrightarrow{V,L} s_3$, де $s_i \xrightarrow{V,L} s_{i+1}$ означає можливість видозміненого ЖЦ, коли, наприклад, етап потрапляння ТП в КС здійснюється не мережею або етап активізації виконується шляхом надходження сигналу мережею, а не локально.

На основі поведінкової моделі побудовано моделі ТП кожного класу з урахування їх особливостей та функціонального навантаження. Так модель класу Trojan-Backdoor матиме вигляд

$$M_{\Theta_{BD}} = \langle \Theta_{BD}, A_{BD}, B_{BD}, W_\nu, Inf, X, Y, Z \rangle, \quad (2)$$

де Θ_{BD} — множина троянських програм класу Trojan-Backdoor; $A_{\Theta_{BD}} = A'_{\Theta_{BD}} \cup A''_{\Theta_{BD}}$ — дії ТП; $A'_{\Theta_{BD}}$ — дії ТП, з потраплянням якої відбувається створення нового файлу; $A''_{\Theta_{BD}}$ — дії ТП, з потраплянням якої відбувається підміна системних файлів троянськими версіями; $W_\nu \in W$ — множини відправлених з КС файлів, шляхом виконання дій $a \in A$, що утворює множину ознак невірного функціонування структурних одиниць ОС КС $b_{BDi} \in B_{BD}$, $B_{BD} = \{b_{BD1}, b_{BD2}, \ldots, b_{BDn}\}$; Inf — ознака інфікування КС; X — відношення, що описує виконання ТП $\nu \in \Theta$ дій $a \in A$, $(\nu, a) \in X$, де $X \subset \Theta \times A$; Y — відношення дій ТП $a \in A$ та структурних одиниць ОС $(a, b) \in Y$, де $Y \subset A \times B$; Z — відношенням дій ТП $a \in A$ та файлів $w \in W$, $(a, w) \in Z$, де $Z \subset A \times W$.

Діагностування КС на наявність ТП

Процес діагностування КС на наявність ТП складається з двох підпроцесів діагностування в режимах монітора та сканера. Частини процесу діагностування КС на наявність ТП позначено як Ω, $\Omega = \{\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_6\}$ та Δ, $\Delta = \{\Delta_1, \Delta_2, \Delta_3, \Delta_4\}$. Процес діагностування в режимі монітора складається з етапів: Ω_1 — відслідковування потоків, що здійснюються через системні порти КС; Ω_2 — відслідковування виконання системних функцій в КС; Ω_3 — блокування виконання програмним об'єктом системних функцій, підозрілість яких визначена на інших етапах процесу АД; Ω_4 — фазифікація в межах системи нечіткого логічного висновку (НЛВ) для введення нечіткості шляхом задання ступенів підозрілості функціонування ПЗ та ступенів небезпеки інфікування КС; Ω_5 — робота машини логічного висновку в межах системи НЛВ; Ω_6 — виконання процедури дефазифікації в межах системи НЛВ для визначення ступеня небезпеки інфікування КС троянською програмою. Процес сканування включає такі етапи: Δ_1 — формування набору файлів, що

підлягають процедурі створення набору захищених бінарних послідовностей; Δ_2 — генерація набору шаблонів файлів, відібраних на попередньому етапі та виконання кодування даних у визначеному форматі; Δ_3 — генерація детекторів згідно з обраним алгоритмом; Δ_4 — сканування КС зіставлення захищених двійковий послідовностей об'єктів антивірусного діагностування зі згенерованими на попередньому етапі детекторами. З урахуванням зворотних зв'язків між етапами покажемо на рис. 1 схему процесу діагностування КС на наявність ТП.

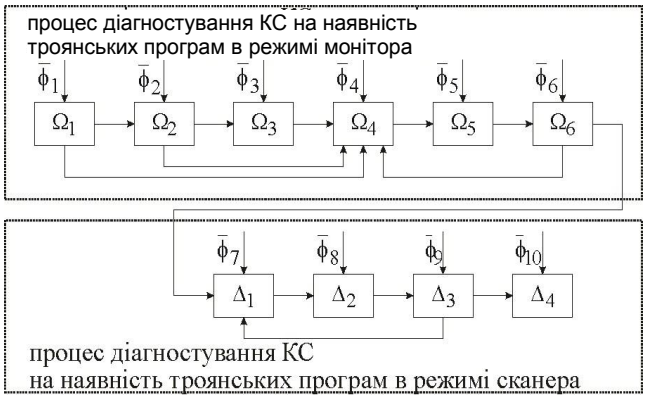

Рис. 1. Формалізована схема процесу діагностування КС на наявність ТП

Для формалізації виконання етапів АД розроблено модель процесу діагностування КС на наявність ТП [2] з урахуванням параметрів, які використовують вищевказані етапи у вигляді

$$M_{\upsilon} = \langle \{E, R, M_W, f_m\}, \{E, H, S, D, E_v, f_s\} \rangle, \tag{4}$$

де для етапів $\Omega_1 - \Omega_6$: E — множина об'єктів діагностування в режимі монітора $e_k \in E$, а саме множина файлів КС, причому $\Theta \in E$; R — підсумкове число $R \in [0,1]$, яке свідчить про ступінь небезпеки інфікування КС троянською програмою; відношення ε між об'єктами та станами, при чому для $\nu \in \Theta$ та $s \in S$; $f_m(I_m, I'_m, I''_m)$ — функція адаптивності діагностування КС в режимі монітора, параметри якої змінюються в залежності від вхідних даних, де I_m — набір діагностичної інформації, $I_m = \langle \Theta, V, L, R, \rangle$); I'_m — вектор результатів антивірусного діагностування, $I'_m = \langle R_1, R_2, \ldots, R_n \rangle$; I''_m — набір даних про виявлене ШПЗ, які збираються і використовуються в майбутньому як знання, $I''_m = \langle E, R \rangle$; для етапів $\Delta_1 - \Delta_4$: E — множина об'єктів діагностування в режимі сканера $e_k \in E$, H — множина об'єктів $h \in H$, що підлягають процедурі сканування на предмет можливого факту їх підміни; S — множина захищених двійкових послідовностей $s \in S$; D — множина детекторів, згенерованих для сканування системи $d \in D$, E_v — множина файлів КС, що були підмінені троянськими версіями; $f_s(I_s, I'_s, I''_s)$ — функція адаптивності діагностування КС в режимі сканера, де I_s — набір діагностичної інформації, $I_s = \langle H, S, D \rangle$; I'_s — результати антивірусного сканування подані набором файлів, що були підмінені троянськими версіями, $I'_s = \langle E_1, E_2, \ldots, E_n \rangle$; I''_s — вектор інформації про оновлення системних файлів та встановлення нового ПЗ, як компонентів об'єкта діагностування $I''_s = \langle E'_1, E'_2, \ldots, E'_n \rangle$.

Адаптивна ІТ діагностування комп'ютерних систем на наявність троянських програм

Розроблено адаптивну ІТ діагностування КС на наявність ТП, яка дозволяє здійснити висновок щодо можливого інфікування КС троянською програмою як відомою, так і новою, а також дозволяє виявляти факт підміни системних файлів троянськими версіями [3]. Процес діагностування КС на наявність ТП адаптується налаштуванням його параметрів в залежності від особливостей КС, що діагностується, а саме: типу операційної системи, встановленого ПЗ в КС, поведінки ТП, накопичених в процесі експлуатації КС. АІТ діагностування КС на наявність ТП подано схемою на рис. 2.

Розроблена адаптивна ІТ діагностування КС на наявність ТП включає в себе метод діагностування КС на наявність ТП в режимі монітора та метод побудови захищених послідовностей та генерації детекторів для діагностування КС на наявність ТП в режимі сканера. Процес діагностування КС на наявність ТП подано схемою на рис. 3.

Рис. 2. Адаптивна інформаційна технологія діагностування комп'ютерних систем на наявність ТП

Метод діагностування КС на наявність ТП в режимі монітора

Розроблено новий метод діагностування КС на наявність ТП в режимі монітора, який полягає в застосуванні апарату нечіткої логіки, і дає можливість зробити висновок щодо ступеня небезпеки інфікування КС троянською програмою [4]. З цією метою здійснюється побудова вхідної та вихідної лінгвістичних змінних з іменами: «Ступінь підозрілості програмного об'єкта» — для вхідної лінгвістичної змінної, і «Ступінь небезпеки інфікування» — для вихідної.

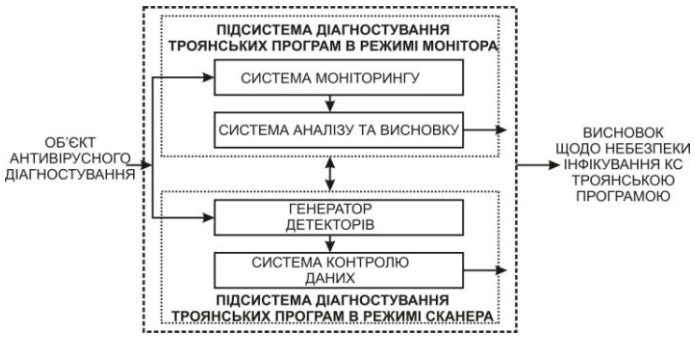

Рис. 3. Схема процесу діагностування КС на наявність троянських програм

Для формування функції належності для вхідної лінгвістичної змінної розроблено новий метод, суть якого полягає в знаходженні для кожної дії найімовірнішого порту потрапляння шляхом ранжування з побудовою матриці переваги $S = \left|s_{ij}\right|$, де $s_{ji} = 1/s_{ij}$;

$s_{ij} = \sum_{k=1}^{r} s_{ij}^{k} \cdot p_k \Big/ \sum_{k=1}^{r} s_{jk}^{k} \cdot p_k$; $s_{ii} = 1;\ i, j = \overline{1, m};$, $s_{ij} = s_i / s_j$, $0 < s_{ij} < \infty$. Потім знаходиться для матриці S власний вектор $\Pi = (\pi_1, \ldots, \pi_m)$, що відповідає максимальному додатному кореню λ характеристичного полінома $|S - \lambda E| = 0$; $S\Pi = \lambda\Pi$, де E — одинична матриця. Компоненти вектора $\Pi\left(\sum \pi_i = 1\right)$ ототожнюються з оцінкою $\mu_{X^p}(x_i, y_j)$. У результаті отримуємо матрицю відношення $V_p = \left|x_i, y_j\right|$, у якій кожному відношенню (x_i, y_j) відповідає значення $0 \le \pi \le 1$. Наступним кроком методу є побудова оптимізованої матриці $V_p^* = \left|x_i, y_j\right|$, з відношень (x_i, y_j) зі значеннями $\pi_{\max}$ $(0 \le \pi_{\max} \le 1)$ та побудова нормованої кривої функції належності $\mu_{X^p}(R)$ вхідної змінної. Для розв'язання поставленої задачі було реалізовано систему НЛВ з використанням алгоритму Мамдані. Графічне зображення результатів системи НЛВ подано на рис. 4.

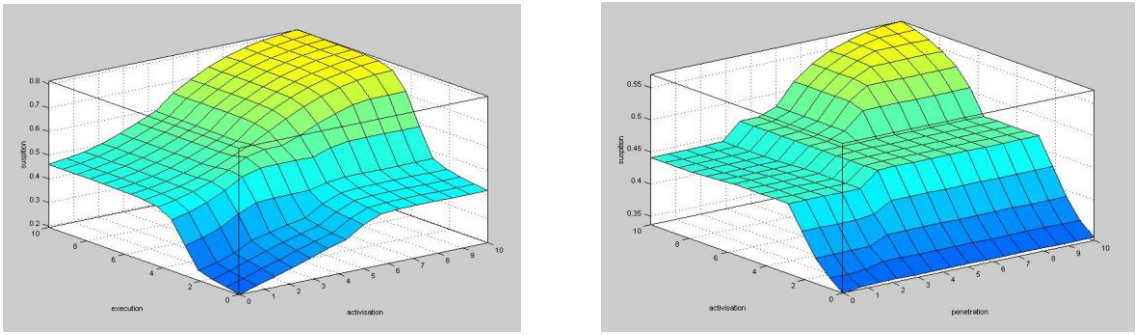

Рис. 4. Результат нечіткого логічного висновку

Метод також передбачає можливість накопичення нових поведінок ТП як знань для підвищення достовірності діагностування КС на наявність ТП.

Метод побудови захищених послідовностей для діагностування КС на наявність ТП в режимі сканера

Розроблено новий метод побудови захищених послідовностей та генерації детекторів для діагностування КС на наявність ТП в режимі сканера, в основі якого лежить застосування алгоритмів штучних імунних систем [5], і який дозволяє виявляти факт підміни системних фалів троянськими версіями в залежності від типу ОС та ПЗ окремо взятої КС. Метод передбачає виконання таких кроків:

1) формування набору файлів для сканування: системних бібліотек, виконуваних файлів системних служб та драйверів пристроїв КС, які можна вважати еталонними;

2) генерація захищених двійкових послідовностей згідно з форматом для операційної системи КС. Для операційної системи типу GNU/Linux захищена послідовність має формат

$$D_i^L = \langle m_1 \ldots m_i \ldots m_{x1}, u_1 \ldots u_i \ldots u_{x2}, g_1 \ldots g_i \ldots g_{x3}, s_1 \ldots s_i \ldots s_{x4}, t_1 \ldots t_i \ldots t_{x5}, C_1 \ldots C_i \ldots C_{x6} \rangle, \quad (5)$$

де $m_1 \ldots m_i \ldots m_{x1}$ — режим файлу (тип і права доступу); $u_1 \ldots u_i \ldots u_{x2}$ — числовий ідентифікатор власника файлу, який показує власника файлу; $g_1 \ldots g_i \ldots g_{x3}$ — числовий ідентифікатор групи власника файлу; $s_1 \ldots s_i \ldots s_{x4}$ — розмір файлу; $t_1 \ldots t_i \ldots t_{x5}$ — час останньої зміни файлу; $C_1 \ldots C_i \ldots C_{x6}$ — CRC файлу, якщо $i = \overline{1,n}$, де n — кількість детекторів.

Детектори для ОС типу MS Windows генеруватимемо у форматі

$$D_i^W = \langle s_1 \ldots s_i \ldots s_{z1}, t_1 \ldots t_i \ldots t_{z2}, a_1 \ldots a_i \ldots a_{z3}, C_1 \ldots C_i \ldots C_{z4} \rangle, \quad (6)$$

де $s_1 \ldots s_i \ldots s_{z1}$ — розмір файлу; $t_1 \ldots t_i \ldots t_{z2}$ — час останньої зміни файлу; $a_1 \ldots a_i \ldots a_{z3}$ — атрибут файлу (параметри: лише читання, прихований, системний, архівний); $C_1 \ldots C_i \ldots C_{z4}$ — CRC файлу, для $i = \overline{1,n}$, де n — кількість детекторів;

3) на етапі антивірусного сканування КС виконання зіставлення захищених двійкових послідовностей з детекторами;

4) у випадку збігу захищених послідовностей з детектором виконання сповіщення про виявлення підміни та перевірка на підозрілість поведінки програмного об'єкту.

Для запропонованої ІТ розроблені алгоритми діагностування КС на наявність ТП [6]. Складність алгоритму діагностування КС на наявність ТП в режимі монітора, який включає пошук схожої поведінки досліджуваного програмного об'єкта з поведінкою ТП в базі, складає $O\left((l+l')n\right)$, де l' — довжина найбільшої спільної послідовності ω серед k послідовностей ξ довжини l, $\|\omega^*\| = \max(\{\|\omega\| \mid \omega \subseteq \xi_i,\ i = \overline{1,k}\})$. Складність алгоритму діагностування КС на наявність ТП в режимі сканера в на етапі підготовки даних до сканування складає: $O\left((l-r)2^r \cdot N_s\right) + O\left((l-r)2^r \cdot N_R\right) + O\left(N_s \log_2 N_s\right)$, де $O\left((l-r)\cdot 2^r \cdot N_s\right)$ — час на генерацію двійкових послідовностей, $O\left((l-r)\cdot 2^r \cdot N_R\right)$ — час на генерацію детекторів, $O\left(N_s \log_2 N_s\right)$ — час сортування двійкових послідовностей, N_R — число двійкових послідовностей-детекторів після їх перевірки на збіг із захищеними послідовностями; l — довжина послідовності, r — число розрядів послідовності збігу, N_S — число захищених послідовностей. На етапі діагностування КС в режимі сканера здійснюється порівняння захищених послідовностей із детекторами складає $O\left(N_s \log_2 N_R\right)$.

Розроблено програмне забезпечення (ПЗ), що реалізує адаптивну інформаційну технологію діагностування КС на наявність ТП [7]. В процесі експлуатації СДКС відбувається автоматичне накопичення інформації про виявлене ШПЗ, занесення його до бази поведінок (для антивірусного монітора), та автоматична генерація набору захищених послідовностей та детекторів у випадку оновлення чи встановлення нового ПЗ (для антивірусного сканера). Порівняльний аналіз достовірності та ефективності діагностування роботи відомих АЗ із розробленим ПЗ, що ґрунтується на використанні адаптивної ІТ, подано гістограмою на рис. 5 та табл. 1 [6].

Отримані результати діагностування КС на наявність ТП показали приріст достовірності 5—15 % та підвищення ефективності у 1,4 рази у порівнянні з відомими засобами діагностування КС на наявність нових ТП.

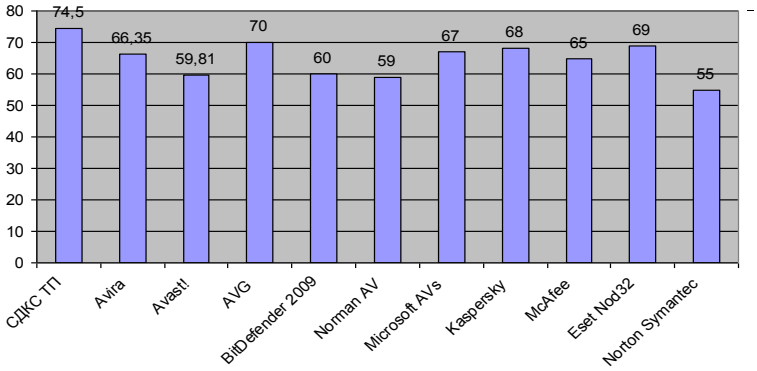

Рис. 5. Достовірність діагностування КС у порівнянні з відомими технологіями

Висновки

Розроблено поведінкову модель ТП та поведінкові моделі класів ТП урахування особливостей функціонування протягом їх життєвого циклу та деструктивного характеру дій у КС, що уможливлює підвищити достовірність їх виявлення в КС. Розроблено модель процесу діагностування КС на наявність ТП, яка базується на залученні компонентів штучного інтелекту, зокрема нечіткої логіки та алгоритмів штучних імунних систем, і відрізняється від відомих тим, що використовує поведінкові моделі класів ТП, дозволяє адаптувати процес діагностування до окремо взятої КС і не потребує побудови баз сигнатур.

Розроблено метод діагностування КС на наявність ТП в режимі монітора на основі нечіткого логічного висновку з оцінкою лінгвістичної змінної на базі попарного порівняння експертом з урахуванням оціночних ознак, яке дозволяє врахувати особливості порівнюваних об'єктів, і яке не вимагає виконання умови транзитивності, що дає можливість визначити ступінь небезпеки інфікування комп'ютерних систем троянськими програмами в умовах апріорної невизначеності. Розроблено метод побудови захищених послідовностей та генерації детекторів на основі штучних імунних систем, який уможливлює визначення факту підміни файлів троянськими версіями в процесі діагностування КС на наявність ТП в режимі сканера.

Розроблено адаптивну інформаційну технологію діагностування КС на наявність ТП, суть якої полягає у аналізі поведінки програмного об'єкта в КС та виявленні факту підміни системних файлів троянськими версіями, відмінністю якої від відомих є те, що параметри діагностування автоматично налаштовуються в залежності від особливостей КС. Це дає можливість виявляти нові ТП та підвищити достовірність та ефективність процесу діагностування КС на наявність ТП. Розроблені алгоритми діагностування КС на наявність ТП. Дослідження їх складності показало можливість їх програмної реалізації в межах адаптивної інформаційної технології. Розроблено ПЗ, що реалізує адаптивну інформаційну технологію діагностування КС на наявність ТП, і задачею якого є діагностування КС на наявність як відомих так нових ТП. Отримані результати досліджень показали підвищення достовірності діагностування на 5—15 %, та ефективності у 1,4 рази у порівнянні з відомими засобами діагностування КС на наявність нових ТП.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Касперський Е. В. Компьютерное зловредство. — 1-е изд. / Е. В. Касперський. — СПб. : Питер, 2009. — 208 с.

2. Савенко О. С. Модель процесу пошуку троянських програм в персональному комп'ютері / Олег Савенко, Сергій Лисенко // Радіоелектронні і комп'ютерні системи. — 2008. — № 7. — С. 87—92.

3. Савенко О. Інтелектуальний метод та алгоритми пошуку троянських програм в персональних комп'ютерах / О. С. Савенко, С. М. Лисенко // Вісник Вінницького політехнічного інституту. — 2008. — № 6. — С. 129—137.

4. Графов Р. П. Использование нечеткой логики для поиска троянских программных продуктов в вычислительных системах / Р. П. Графов, О. С. Савенко, С. М. Лысенко // Вісник Чернівецького національного університету. — 2009. — № 6. — С. 85—91.

5. Савенко О. С. Розробка процесу виявлення троянських програм на основі використання штучних імунних систем / О. С. Савенко, С. М. Лисенко // Вісник Хмельницького національного університету. — 2008. — № 5. — С. 183—188.

6. Савенко О. С. Алгоритми пошуку троянських програм в персональних комп'ютерах / О. С. Савенко, С. М. Лисенко // Радіоелектронні і комп'ютерні системи. — 2009. — № 5. — С. 120—126.

7. Лисенко С. М. Розробка програмного забезпечення реалізації інтелектуального методу пошуку троянських програм в персональних комп'ютерах / С. М. Лисенко, А. П. Гонтар, А. С. Шевцов // Вісник Хмельницького національного університету. — 2010. — Том 1, № 1. — С. 98—105.

Рекомендована кафедрою комп'ютерних систем управління

Стаття надійшла до редакції 11.02.11
Рекомендована до друку 8.07.11

Савенко Олег Станіславович — доцент, ***Лисенко Сергій Миколайович*** — старший викладач.
Кафедра системного програмування, Хмельницький національний університет, Хмельницький