

МЕТОДИКА СИНТЕЗУ АЛГОРИТМІВ РОЗПІЗНАВАННЯ УРАЗЛИВОСТЕЙ WEB-РЕСУРСІВ ЗА СИГНАТУРАМИ ЗНАЧЕНЬ НЕЧІТКИХ ЛІНГВІСТИЧНИХ ОЗНАК

Анотація. Формалізовано задачу нечіткого розпізнавання уразливостей web-ресурсів, що задана складним еталонним описом у вигляді сигнатур інтервалів значень нечітких лінгвістичних ознак. Із застосуванням отриманого еталонного опису та математичного апарату теорії перевірки складних статистичних гіпотез синтезовано алгоритми багатоальтернативного розпізнавання об'єктів тестування за мінімаксімним вирішувальним правилом, а також критеріями баєсівського максимуму апостеріорної ймовірності та максимальної правдоподібності.

Ключові слова: нечіткі лінгвістичні ознаки, розпізнавання, еталонний опис, сигнатура.

ВСТУП

Характерною ознакою теперішнього часу є розвиток програмних застосувань інформаційного середовища телекомунікаційних систем (ІС ТКС) з вбудованими автоматизованими процедурами розпізнавання уразливостей програмного забезпечення web-ресурсів для запобігання несанкційного доступу до інформації.

Найбільш складними ключовими задачами, які розв'язують при автоматизованому розпізнаванні уразливостей web-ресурсів (можливих сценаріїв атак для несанкційного проникнення), є формування робочого словника інформаційних ознак, сигнатур, формалізація задачі та синтез алгоритмів ухвалення рішення про відповідність одному із заданих образів вибірки значень інформаційних сигнатур.

У залежності від режиму роботи комплексу тестування ІС ТКС при розпізнаванні застосовують методи перевірки складних статистичних гіпотез.

Розв'язання задач розпізнавання у комплексах тестування ІС ТКС ускладнено труднощами формалізації сигнатур та ознак, їхньою неоднорідною структурою, наявністю перетинних класів образів. Це зумовлено тим, що на вхід пристрою первинного оброблення та формування інформаційних сигнатур надходять результати застосування тестових впливів, які структурно являють собою стандартні для теорії нечіткої логіки лінгвістичні вислови оператора комплексу: «правильно», «неправильно», «абсолютно правильно», «абсолютно неправильно». Кожне значення лінгвістичної змінної є нечіткою множиною одиничного інтервалу. На основі поняття лінгвістичної змінної, запропонованого Л. Заде [1, 2], кожному тестовому впливу з множини сигнатур ознак ставиться у відповідність проста лінгвістична змінна, а кожній сигнатурі — складна. Однак у відомій літературі питання формалізації еталонного опису інформаційних сигнатур та ознак, синтезу нечітких алгоритмів розпізнавання уразливості web-ресурсів за реалізацією сукупності тестових впливів — лінгвістичних висловлювань оператора в апаратно-програмних комплексах тестування ІС ТКС не розв'язані у повному обсязі.

На цей час в ІС ТКС для тестування web-ресурсів на наявність уразливостей оператори використовують спеціальне програмне забезпечення (СПЗ) [3]. Але навіть за допомогою СПЗ процес пошуку уразливостей не має єдиного еталонного опису та алгоритму для розпізнавання уразливостей web-ресурсів.

У роботі [4] описано приклади тестування web-ресурсу на наявність уразливостей та спробу узагальнення цього процесу. Реалізацію процесу тестування web-ресурсу на наявність уразливостей за допомогою пошукової системи Google наведено в [5]. У статті [6] формалізовано складний еталонний опис для розпізнавання на основі введення ймовірностей належності образу кожному з класів образів та їхніх ймовірнісно-статистичних моделей. Однак нечіткий характер інформаційних ознак та сигнатур, недостатність обсягу навчальної вибірки суттєво ускладнюють розроблення еталонного опису із застосуванням ймовірнісно-статистичних моделей визначення стану безпеки web-ресурсів [7].

З огляду на це дослідження логічних та нечітких методів і підходів є актуальними та мають практичне значення.

ПОСТАНОВКА ЗАДАЧІ

Метою статті є класифікація уразливостей web-ресурсів та розроблення формалізованого складного еталонного опису ознак і максимально автоматизованого нечіткого алгоритму розпізнавання уразливостей в комплексах тестування web-ресурсів.

Задачу класифікації уразливостей (способів порушення безпеки) web-ресурсів можна формалізувати як модифікацію задачі розпізнавання образів за вибіркою лінгвістичних змінних.

Нехай на виході пристрою первинного оброблення та визначення нечітких моніторингових сигнатур комплексу тестування ІС ТКС фіксуються вхідні спостереження-відгуки, що являють собою реакцію web-ресурсу на сукупність тестових впливів. Результати тестових впливів визначає оператор у вигляді логічних нечітких змінних (ЛгА): «строго неможливо» (СНМ), «неможливо» (НМ), «скоріш за все можливо» (СЗМ), «можливо» (М) та «строго можливо» (СМ).

На універсумі Ψ об'єктів розпізнавання визначено K образів Ψ_i , $i \in \{1, 2, \dots, K\}$, об'єктів тестування або стану уразливості web-ресурсу.

За підходом [1] сформовано оптимальний за визначеним показником (наприклад, мінімальна розмірність при забезпеченні потрібного рівня інформативності) робочий словник ознак S та на його основі обчислено множину моніторингових сигнатур \mathbf{X}^S . Кожна з сигнатур містить еквівалентні значення \mathfrak{S} нечітких моніторингових ознак $s_j \in S$, $j \in \{1, 2, \dots, \mathfrak{S}\}$. Кожний з K образів характеризують відповідною сигнатурою (вектором) $\mathbf{x}_i^S = \{x'_{i1} \dots x''_{i1}, x'_{i2} \dots x''_{i2}, \dots, x'_{i\mathfrak{S}} \dots x''_{i\mathfrak{S}}\}$, $i = 1, \dots, K$, $\mathbf{x}_i^S \in \mathbf{X}^S$, $\mathbf{X}^S \subset R^n$, яка, по суті, являє собою впорядковані сукупності інтервалів нечітких значень складних лінгвістичних ознак, що визначені у цифровій формі. Кожному значенню лінгвістичної ознаки у моніторинговій сигнатурі Ψ_i образу ставиться у відповідність інтервал значень (за кількістю класів) з використанням функцій належності μ_i (рис. 1).

На основі тестування web-ресурсу оператор формує вибірку визначених еквівалентних значень $x = (x_1, x_2, \dots, x_{\mathfrak{S}})$ лінгвістичних змінних, що є результатом впорядкованої оцінки значень нечітких ознак s_j , $j \in \{1, 2, \dots, \mathfrak{S}\}$.

За відсутності випадкових факторів та помилок оператор при позитивному виконанні тестового впливу «запит – відгук» вирішує, що значення лінгвістичної змінної ототожнюється з логічною одиницею (див. рис. 1), а в іншому випадку — з логічним нулем. Однак на практиці на коректність інтерпретації результатів тестового запиту впливає складність визначення меж відповідності лінгвістичних змінних та сукупність випадкових факторів. Тому еталонний опис тестових сигнатур являє собою сукупності інтервалів значень s_j ознак з діапазоном від $0 + j - 1$ до $1 + j - 1$ (див. рис. 1), що задані у вигляді сукупності відповідних умовних щільнос-

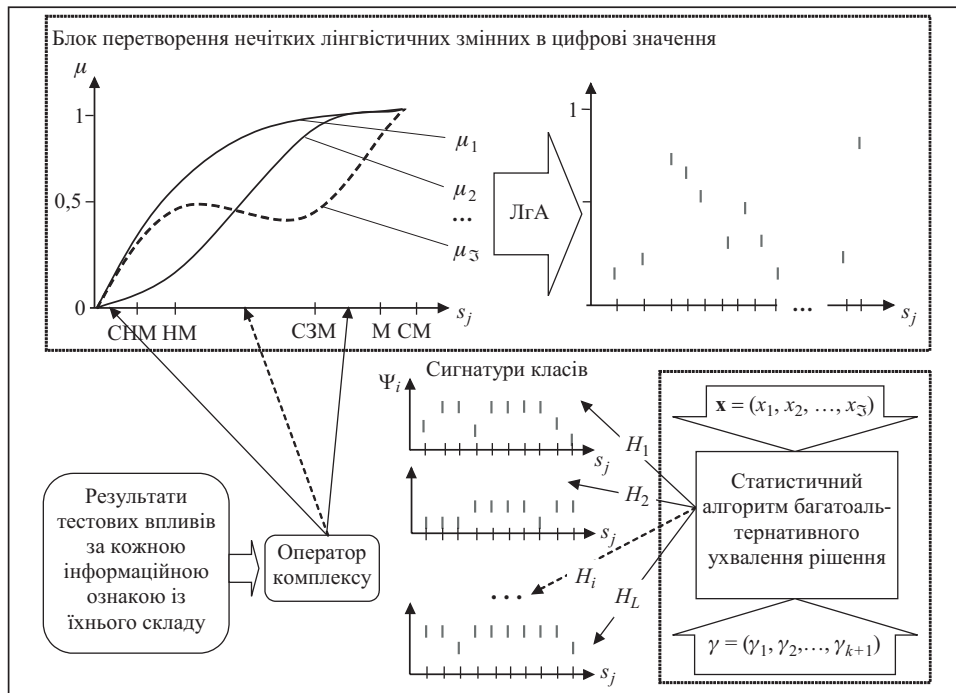


Рис. 1. Схема формалізації задачі автоматизованого розпізнавання уразливостей web-ресурсів за нечіткими сигнатурами ознак в комплексах тестування

тей ймовірностей значень ознак $w_{jr}(x_j, x'_{jr}, x''_{jr})$. Ці щільності характеризують еталонний розподіл значень x_j j -ї ознаки s_j на кожному r -му інтервалі можливих значень x'_{jr}, \dots, x''_{jr} для кожного з образів, що розпізнаються.

Під час застосування алгоритмів розпізнавання перевіряють гіпотези H_i про те, що вибірка впорядкованих значень лінгвістичних нечітких змінних відповідає сигнатурі $\mathbf{x}_i^s = \{x'_{i1} \dots x''_{i1}, x'_{i2} \dots x''_{i2}, \dots, x'_{i3} \dots x''_{i3}\}$, $i = 1, \dots, K$, яка належить образу Ψ_i . У результаті перевірки повинно бути ухвалене одне з L рішень γ_i , що відповідає коректності гіпотези H_i .

Необхідно визначити нерандомізоване правило δ , яке дозволяє згідно з критерієм оптимальності розробити алгоритм, що функціонує відповідно до результатів визначення лінгвістичної ознаки s на L неперетинних областях x_i , $i = 1 \dots 3$ (див. рис. 1). Якщо у процесі розпізнавання значення вибірки $\mathbf{x} = (x_1, x_2, \dots, x_3)$ потраплять одночасно у відповідні області x_j , $j \in \{1, 2, \dots, 3\}$, сигнатури, то ухвалюють рішення γ_i .

В евристичних алгоритмах розпізнавання неперервні інтервали описують множинами дискретних значень. Після цього отримують результативну дискретну множину еталонних значень ознак, яка складається з дискретних значень і перетворених у дискретну форму неперервних інтервалів. Потім перевіряють множину простих гіпотез про відповідність спостережуваної вибірки кожному з дискретних еталонних значень ознак. Для цього використовують множину часткових алгоритмів розпізнавання. За отриманими проміжними результатами розпізнавання остаточно вирішують, чи належить вибірка до одного із образів.

При виконанні таких алгоритмів велика кількість операцій призводить до зайвих витрат часу на формування рішення. На заключному етапі розпізнавання часто застосовують евристичні алгоритми ухвалення остаточних рішень (голосування). При цьому отримують великі значення ймовірностей помилкових рішень, оцінити котрі апріорі, зазвичай, складно. Крім того, на практиці зустрічаються

випадки, коли дискретні еталонні значення ознак нерівноймовірні і (або) у неперервних інтервалах еталонний розподіл ознаки не можна описати рівномірним законом. Для спрощення у таких випадках іноді часткові граничні значення вважають однаковими. Це призводить до додаткових помилок в ухваленні рішень.

Уникнути перерахованих труднощів у побудові алгоритмів розпізнавання можна, застосувавши методи перевірки складних статистичних гіпотез [7, 8]. При цьому громіздкість теоретичних викладок компенсують досягнутим позитивним ефектом — однокроковим алгоритмом, оптимальним за одним із статистичних критеріїв. Крім того, ці алгоритми дозволяють мінімізувати помилки, які виникають через параметричну невизначеність щодо дійсних значень ознак спостережуваного образу.

МЕТОДИКА СИНТЕЗУ СТАТИСТИЧНИХ АЛГОРИТМІВ РОЗПІЗНАВАННЯ ОБРАЗІВ

Нехай образи $\Psi_i \subset \Psi$, що розпізнаються, задані своїм еталонним описом в області x_i^s \mathfrak{I} -вимірному евклідовому простору еталонів \mathbf{X}^s із координатними осями $x_1, x_2, \dots, x_j, \dots, x_{\mathfrak{I}}$. Тоді еталонний опис образів надамо у вигляді суми отриманих у результаті навчання та з'єднаних в інформаційні сигнатури апріорних щільностей ймовірностей $w_j(x)$ еталонних значень ознак $x_1 \dots x$ на множині Ψ_i

$$w_j(x) = \sum_{r=1}^{R_j} p_{jr} w_{ir}(x_j^e, x_{jr}^{e'}, x_{jr}^{e''}), \quad \sum_{r=1}^{R_j} p_{jr} = 1. \quad (1)$$

Значення вхідних параметрів системи — $R_{(j)}$ зважених щільностей ймовірності $w_{jr}(x_j^e, x_{jr}^{e'}, x_{jr}^{e''})$ еталонного розподілу ознаки x_j^e на кожному r -му інтервалі можливих значень $x_{jr}^{e'}, \dots, x_{jr}^{e''}$ j -ї ознаки для кожного образу, що розпізнається, де p_{jr} — умовні апріорні ймовірності потрапляння значення ознаки в r -й інтервал, $r \in \{1, 2, \dots, R_{(j)}\}$, j -ї ознаки в i -му образі, $R_{ij} = \nu_i$, $R_{qj} = \nu_q$ $\forall j \in \{1, 2, \dots, \mathfrak{I}\}$.

Один еталонний інтервал значення j -ї ознаки x_j^e , $j = \{1, 2, \dots, \mathfrak{I}\}$, в одному образі відповідає одному тестовому впливу в координаті цієї ознаки, а кожна сигнатура відповідно містить \mathfrak{I} інтервалів значень ознак. Область визначення функцій щільності ймовірностей $w_{jr}(x_j^e, x_{jr}^{e'}, x_{jr}^{e''})$ значень складного еталонного опису (1) наведемо у вигляді

$$w_{jr}(x_j^e, x_{jr}^{e'}, x_{jr}^{e''}) = \begin{cases} 0 & \text{для } x_j^e < 0 + j - 1, \\ 0 \dots 1 & \text{для } x_{jr}^{e'} \leq x_j^e \leq x_{jr}^{e''}, \\ 0 & \text{для } x_j^e > 1 + j - 1. \end{cases} \quad (2)$$

Висувається L гіпотез H_1, H_2, \dots, H_L про те, що спостережувана вибірка \mathbf{x} (розміром $\zeta \times \mathfrak{I}$) ζ -кратно оцінених значень \mathfrak{I} ознак належить одному з описаних образів Ψ_i (відповідає значенням сигнатури ознак відповідного класу). Простір рішень складається з L елементів γ_i — рішень про ухвалення гіпотези H_i .

Задача полягає в тому, щоб вибрати нерандомізоване дискретно-аналогове правило — (δ) , котре реалізує поділ простору \mathbf{X} на L перетинних областей x_i ,

$\bigcup_{i=1}^L x_i = \mathbf{X}$. Якщо в процесі розпізнавання відбувається подія, яка полягає в одно-

часному потраплянні кожного значення ознаки з вибірки $\mathbf{x} = (x_1, x_2, \dots, x_{\mathfrak{Z}})$ у відповідні інтервали сигнатури

$$\mathbf{x}_i^{\mathfrak{S}} = \{x'_{i1} \dots x''_{i1}, x'_{i2} \dots x''_{i2}, \dots, x'_{i\mathfrak{Z}} \dots x''_{i\mathfrak{Z}}\}, \quad i = 1, \dots, K, \quad (3)$$

то ухвалюють рішення γ_i .

Розглянемо синтез мінімаксних алгоритмів розпізнавання web-ресурсів — об'єктів тестування, які задані сигнатурами у вигляді складного еталонного опису типу (1), (2).

Неважко показати, що постановка задачі багатоальтернативного розпізнавання легко спрощується до одноальтернативного випадку шляхом розгляду множини Ψ об'єктів розпізнавання, на якій задано два образи: $\Psi_o \subset \Psi$ і $\Psi_{no} \subset \Psi$ (o — об'єкт, но — немає об'єкта). Перевіряються дві гіпотези, H_o і H_{no} , про те, що спостережувана вибірка \mathbf{x} розміром $\zeta \times \mathfrak{Z}$ ζ -кратно вимірюваних значень еквівалентних значень \mathfrak{Z} лінгвістичних ознак s_j відповідає сигнатурі образу Ψ_o або Ψ_{no} . У результаті перевірки необхідно ухвалити одне з двох рішень — γ_o чи γ_{no} , що відповідають коректності гіпотез H_o або H_{no} .

Конкретизуємо еталонний розподіл j -ї ознаки s_j , $j \in \{1, 2, \dots, \mathfrak{Z}\}$ (1), (2) у вигляді суми умовних розподілів $w_{oj}(x_j^e) = W(x_j^e | \Psi_o)$, $w_{noj}(x_j^e) = W(x_j^e | \Psi_{no})$, зважених із невідомими апіорними ймовірностями p_o і $p_{no} = 1 - p_o$ спостереження образів S_o і S_{no} [3, 4]:

$$w_{ij}(x_j^e) = p_o w_{oj}(x_j^e) + p_{no} w_{noj}(x_j^e). \quad (4)$$

Кожну з вхідних щільностей ймовірності $w_{(\cdot)j}(x_j^e)$ у (1) представимо зваженими сумами

$$\begin{aligned} w_{ij}(x_j^e) &= \sum_{r=1}^{R_{ij}} p_{ojr} w_{ojr}(x_j^e, x_{ojr}^{e'}, x_{ojr}^{e''}), \quad \sum_{r=1}^{R_{oj}} p_{ojr} = 1, \\ w_{noj}(x_j^e) &= \sum_{r=1}^{R_{noj}} p_{noj r} w_{noj r}(x_j^e, x_{noj r}^{e'}, x_{noj r}^{e''}), \quad \sum_{r=1}^{R_{noj}} p_{noj r} = 1. \end{aligned} \quad (5)$$

На практиці ймовірності p_{ojr} , $p_{noj r}$ апіорі невідомі, тому однозначний розв'язок задач (1)–(5) одержати складно. Вихід із цієї ситуації полягає в тому, щоб спочатку для кожної ознаки в одному образі вважати ці ймовірності однаковими. Потім у процесі розпізнавання потрібно їх оцінити і скорегувати правило вибору (ухвалення) рішення.

Застосовуючи підхід [5–7], для моделі (1)–(3) можна отримати мінімаксне правило розпізнавання образів Ψ_o і Ψ_{no} , що передбачає порівняння відношення правдоподібності $\Lambda(\mathbf{x})$ з пороговим значенням:

$$\delta_{MM} : \Lambda(\mathbf{x}) \underset{\gamma_i}{\overset{\gamma_q}{>}} \mu_{MM} c^*, \quad c^* = \frac{\Pi_{оно} - \Pi_{оо}}{\Pi_{ноо} - \Pi_{ноно}},$$

де Π_{**} — елементи матриці втрат $\mathbf{\Pi} = \begin{vmatrix} \Pi_{оо} & \Pi_{оно} \\ \Pi_{ноо} & \Pi_{ноно} \end{vmatrix}$, $\Pi_{оно} > \Pi_{оо} \geq 0$,

$\Pi_{qi} > \Pi_{qq} \geq 0$, рядки якої відповідають гіпотезам H_i і H_q , а стовпці — рішенням γ_o і γ_{no} .

Відношення $\mu_{MM} = p_o / p_{no}$ знаходять при ймовірності p_o , що відповідає найбільшому значенню баєсівського ризику. Для цього, використовуючи відо-

мий метод [2], необхідно осереднити за областями \mathbf{X}_o^s і $\mathbf{X}_{но}$ умовні функції ризику $r_o(\mathbf{x}^s)$, $r_{но}(\mathbf{x}^s)$:

$$\int_{\mathbf{X}_o^s} r_o(\mathbf{x}^s) w_o(\mathbf{x}^s) d\mathbf{x}^s = r_o, \quad \int_{\mathbf{X}_{но}} r_{но}(\mathbf{x}^s) w_{но}(\mathbf{x}^s) d\mathbf{x}^s = r_{но},$$

де $r_o(\mathbf{x}^s) = \Pi_{oo}[1 - \alpha(\mathbf{x}^s)] + \Pi_{оно}\alpha(\mathbf{x}^s)$, $r_{но}(\mathbf{x}^s) = \Pi_{ноо}\beta(\mathbf{x}^s) + \Pi_{ноно}[1 - \beta(\mathbf{x}^s)]$;

$$\alpha(\mathbf{x}^s) = \int_{\mathbf{X}_{но}} W(\mathbf{x} | \mathbf{x}^s \in \mathbf{X}_o^s) d\mathbf{x}, \quad \beta(\mathbf{x}^e) = \int_{\mathbf{X}_o} W(\mathbf{x} | \mathbf{x}^s \in \mathbf{X}_{но}^s) d\mathbf{x}$$

— умовні ймовірності помилок першого і другого роду; $W(\mathbf{x} | \mathbf{x}^s)$ — функція правдоподібності вибірки \mathbf{x} при фіксованому векторі \mathbf{x}^s .

Трансцендентному рівнянню $r_o = r_{но}$ відносно μ надамо вигляду

$$\mu_{MM} = \arg \{c^* \alpha(\mu c^*) - \beta(\mu c^*) + c^{**} = 0\}, \quad (6)$$

де $\alpha(\mu c^*)$, $\beta(\mu c^*)$ — повні (осереднені за відповідними щільностями) ймовірності, $w_o(\mathbf{x}^s)$ і $w_{но}(\mathbf{x}^s)$ — умовні ймовірності помилок першого і другого роду, обчислені при пороговому значенні μc^* .

На практиці у випадку незалежних ознак рівняння (3) досить просто записати у явному вигляді. При цьому кожен з еталонних описів $w_o(\mathbf{x}^s)$ і $w_{но}(\mathbf{x}^s)$ являє собою перемноження \mathfrak{Z} відповідних щільностей ймовірності (2) $w_{ij}(x_j^e)$ або $w_{qj}(x_j^e)$, рішення ухвалюється за алгоритмом

$$\delta_{MM} : \int_{\mathbf{X}_{но}^s} W(\mathbf{x} | \mathbf{x}^s) \prod_{j=1}^{\mathfrak{Z}} \left[\sum_{r=1}^{R_{noj}} p_{noj r} w_{noj r}(x_j^e, x_{noj r}^{e'}, x_{noj r}^{e''}) \right] d\mathbf{x}^s \times \dots \\ \dots \times \left\langle \int_{\mathbf{X}_o^s} W(\mathbf{x} | \mathbf{x}^s) \prod_{j=1}^{\mathfrak{Z}} \left[\sum_{r=1}^{R_{ojr}} p_{ojr} w_{ojr}(x_j^e, x_{ojr}^{e'}, x_{ojr}^{e''}) \right] d\mathbf{x}^s \right\rangle^{-1} \begin{matrix} \gamma_q \\ > \\ \mu_{MM} c^* \\ < \\ \gamma_i \end{matrix}, \quad (7)$$

де μ_{MM} , як і раніше, визначається рівнянням (6), у котрому

$$\alpha(\mu c^*) = \int_{\mathbf{X}_o^s} \prod_{j=1}^{\mathfrak{Z}} \left[\sum_{r=1}^{R_{ojr}} p_{ojr} w_{ojr}(x_j^e, x_{ojr}^{e'}, x_{ojr}^{e''}) \right] \int_{\mathbf{X}_{но}(\mu c^*)} W(\mathbf{x} | \mathbf{x}^s) d\mathbf{x} d\mathbf{x}^s, \quad (8)$$

$$\beta(\mu c^*) = \int_{\mathbf{X}_{но}^s} \prod_{j=1}^{\mathfrak{Z}} \left[\sum_{r=1}^{R_{noj}} p_{noj r} w_{noj r}(x_j^e, x_{noj r}^{e'}, x_{noj r}^{e''}) \right] \int_{\mathbf{X}_o(\mu c^*)} W(\mathbf{x} | \mathbf{x}^s) d\mathbf{x} d\mathbf{x}^s, \quad (9)$$

$X_o(\mu c^*)$ і $X_{но}(\mu c^*)$ — допустима і критична щодо гіпотези H_o області \mathbf{X}_o і $\mathbf{X}_{но}$, які визначені за умови $\Lambda(\mathbf{x}) = \mu c^*$.

Якість рішень, ухвалених відповідно до правила (7), можна оцінити повними умовними ймовірностями помилок першого $\alpha(\mu_{MM} c^*)$ і другого $\beta(\mu_{MM} c^*)$ роду. Ці ймовірності визначають формули (8) і (9) за умови $\mu = \mu_{MM}$.

Застосовуючи методи синтезу статистичних алгоритмів [5, 6] для багатоальтернативного випадку, вводимо статистику i -го елемента вектора відношень правдоподібності при перевірці складних гіпотез у вигляді

$$\Lambda_i(\mathbf{x}) = \int_{\mathbf{x}_i^s} W(\mathbf{x} | \mathbf{x}^s) \prod_{j=1}^{\mathfrak{Z}} \left[\sum_{r=1}^{R_{ij}} p_{ijr} w_{ijr}(x_j^e, x_{ijr}^{e'}, x_{ijr}^{e''}) \right] d\mathbf{x}^s \times \left\langle \int_{\mathbf{x}_1^s} W(\mathbf{x} | \mathbf{x}^s) \prod_{j=1}^{\mathfrak{Z}} \left[\sum_{r=1}^{R_{1j}} p_{1jr} w_{1jr}(x_j^e, x_{1jr}^{e'}, x_{1jr}^{e''}) \right] d\mathbf{x}^s \right\rangle^{-1}, \quad i=2, 3, \dots, K. \quad (10)$$

На основі [7] та (10) можна отримати алгоритми, оптимальні щодо основних критеріїв ефективності. Для цього необхідно відповідно до застосованого критерію та умов конкретної задачі визначити відношення правдоподібності (10) і порівняти його з порогом. Зокрема, баєсівський алгоритм багатоальтернативного розпізнавання образів має вигляд

$$\delta_B: \sum_{q=2}^K (\Pi_{qq} - \Pi_{qi}) \frac{p_q}{p_1} \Lambda_q(\mathbf{x}) \geq \Pi_{1i} - \Pi_{1q}, \quad t=1, 2, \dots, K, \quad t \neq i, \quad i=2, 3, \dots, K, \quad (11)$$

де $\Pi_{qi} \geq 0$ — елементи матриці втрат $\mathbf{\Pi}$ розміром $K \times K$.

До області \mathbf{X}_i , $i \in \{2, 3, \dots, K\}$, зараховують точки вибіркового простору \mathbf{X} , що задовільняють систему нерівностей (11). Область \mathbf{X}_1 визначають з умови

$$\mathbf{X}_1 = \mathbf{X} \bigcup_{i=2}^K \mathbf{X}_i.$$

При застосуванні критерію максимуму апостеріорної ймовірності ухвалюють рішення γ_i , $i=2, 3, \dots, K$, якщо

$$\delta: p_i \Lambda_i(\mathbf{x}) = \max_{2 \leq q \leq L} p_q \Lambda_q(\mathbf{x}), \quad (p_q / p_1) \Lambda_q(\mathbf{x}) \geq 1, \quad q=2, 3, \dots, K.$$

При використанні критерію максимальної правдоподібності необхідно серед образів вибрати об'єкти розпізнавання і ухвалити щодо них відповідне рішення. Перевіримо статистики відношень правдоподібності:

$$\Lambda_i(\mathbf{x}) = \frac{\max_{x \in X_i} W(\mathbf{x} | \mathbf{x}^s)}{\max_{x \in X_1} W(\mathbf{x} | \mathbf{x}^s)}, \quad \mathbf{S}_i = \bigcup_{j=1}^{\mathfrak{Z}} \left(\bigcup_{r=1}^{R_{ij}} X_{ijr}, x_{ijd} \right), \quad i=2, 3, \dots, K,$$

де X_{ijr} — r -і інтервали еталонних значень i -го образу в метриці ознаки s_j ; x_{ijd} — d -і еталонні значення i -го образу в метриці j -ї ознаки.

Ухвалюють рішення γ_i , $i=2, 3, \dots, L$, якщо

$$\delta_{МП}: \Lambda_i(\mathbf{x}) = \max_{2 \leq q \leq L} \Lambda_q(\mathbf{x}), \quad \Lambda_q(\mathbf{x}) \geq 1, \quad q=2, 3, \dots, K,$$

і рішення γ_1 , якщо $\Lambda_q(\mathbf{x}) < 1 \quad \forall q=2, 3, \dots, K$.

ВИСНОВКИ

Формалізовано задачу нечіткого розпізнавання об'єктів тестування уразливостей web-ресурсів, що задана складним еталонним описом у вигляді сукупності інтервалів значень нечітких лінгвістичних ознак. Із застосуванням отриманого еталонного опису та математичного апарату теорії перевірки складних статистичних гіпотез запропоновано алгоритми багатоальтернативного розпізнавання

об'єктів тестування за мінімаксним правилом ухвалення рішення, а також баєсівськими критеріями.

Напрямом подальших досліджень є конкретизація наведених вирішувальних правил та оцінка їхньої ефективності для випадків застосування різних моделей щільностей ймовірності $w_{jr}(x_j^e, x_{jr}^{e'}, x_{jr}^{e''})$ еталонного розподілу ознаки x_j^e на кожному r -му інтервалі можливих значень $x_{jr}^{e'}, \dots, x_{jr}^{e''}$ j -ї ознаки для кожного образу, що розпізнається.

СПИСОК ЛІТЕРАТУРИ

1. Беллман Р., Заде Л. Принятие решений в расплывчатых условиях. В кн.: Вопросы анализа и процедуры принятия решений. Москва: Мир, 1976. С. 172–216.
2. Заде Л. Понятие лингвистической переменной приближенных решений. Москва: Мир, 1976. 520 с.
3. Автоматический поиск уязвимостей. Журнал «Хакер». 2010. URL: www.xakep.ru/2006/03/10/30503.
4. Поиск уязвимостей. Пособие для чайников. 2015. URL: www.pikabu.ru/story/poisk_uyazvimostey_posobie_dlya_chaynikov_2961936.
5. Ищем уязвимости при помощи Google. 2016. URL: www.habrahabr.ru/post/283210.
6. Певцов Г.В., Лулундин В.А. Синтез непараметрических алгоритмов распознавания групп радиоизлучений с оценением усредненных функций правдоподобия методом гистограмм. *Системы обработки информации*. 2003. № 3. С. 23–34.
7. Левин Б.Р. Теоретические основы статистической радиотехники. Кн. 2. Москва: Сов. радио, 1975. 392 с.
8. Безрук В.М., Певцов Г.В. Теоретические основы проектирования систем распознавания сигналов для автоматизированного радиоконтроля. Харьков: Коллегиум, 2007. 430 с.

Надійшла до редакції 30.12.2016

А.А. Ильяшов

МЕТОДИКА СИНТЕЗА АЛГОРИТМОВ РАСПОЗНАВАНИЯ УЯЗВИМОСТЕЙ WEB-РЕСУРСОВ ПО СИГНАТУРАМ ЗНАЧЕНИЙ НЕЧЕТКИХ ЛИНГВИСТИЧЕСКИХ ПРИЗНАКОВ

Аннотация. Формализована задача нечеткого распознавания уязвимостей web-ресурсов, заданная сложным эталонным описанием в виде сигнатур интервалов значений нечетких лингвистических признаков. С использованием полученного эталонного описания и математического аппарата теории проверки сложных статистических гипотез синтезированы алгоритмы многоальтернативного распознавания объектов тестирования по минимаксному решающему правилу, а также критериям байесовского максимума апостериорной вероятности и максимальной правдоподобности.

Ключевые слова: нечеткие лингвистические признаки, распознавание, эталонное описание, сигнатура.

О.А. Piashov

SYNTHESIS OF ALGORITHMS FOR RECOGNITION OF VULNERABILITIES IN WEB-RESOURCES THROUGH SIGNATURES OF FUZZY LINGUISTIC FEATURES

Abstract. The paper formalizes the problem of fuzzy recognition of vulnerabilities in web-resources that is set by a complex reference description in the form of signatures of intervals of the values of fuzzy linguistic features. The reference description and the mathematics of the theory of validation complex statistical hypotheses were applied to synthesize multi-criteria object recognition algorithms using the minimax decision rule and the Bayesian maximum of a posteriori probability and maximum likelihood criteria.

Keywords: fuzzy linguistic features, recognition, reference description, signature.

Ильяшов Александр Авксентійович,

доктор військ. наук, доцент, головний науковий співробітник Центрального науково-дослідного інституту Збройних Сил України, Київ, e-mail: aleksandr.ilyashov@gmail.com.