MATHEMATICAL AND COMPUTER MODELING

УДК 004(942+056.5)

ДМИТРО ШАРАДКІН

ВИКОРИСТАННЯ КРИТЕРІЮ ВИЯВЛЕННЯ ЗМІН ПОВЕДІНКИ ОБ'ЄКТА НА ОСНОВІ АНАЛІЗУ КОЕФІЦІЄНТА АВТОКОРЕЛЯЦІЇ В ЗАДАЧАХ ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

Розглядаються методи виявлення факту зміни поведінки об’єктів, зокрема сучасних інформаційно-комп’ютерних мереж, на основі аналізу їх моделей функціювання у вигляді часових рядів. Показано, що вказані об’єкти характеризуються великою внутрішньою складністю а також різноманітністю статистичних законів розподілення їх значень. Найскладнішим для дослідження та практичного застосування є наявність широкого спектру можливих форм та характеристик змін поведінки, які викликані непередбаченістю як самих причин, так і їх можливого впливу на такі об’єкти. Вказані обмеження не дають можливості обрати єдиний універсальний критерій виявлення змін в поведінці об’єктів дослідження, який міг би бути спроможним реагувати на більшість змін, та приводять до необхідності спільного застосування ансамблю відповідних критеріїв. В роботі детально розглядається один з таких критеріїв, а саме критерій, що базується на використання коефіцієнту автокореляції часового ряду першого порядку. За допомогою засобів статистичного моделювання проведений аналіз основних характеристик критерію, вивчені показники його статистичної потужності, ефективності та різноманітні обмеження його використання. Проаналізовані залежності результатів використання критерію від параметрів алгоритму виявлення змін. Проведено порівняння запропонованого критерію з іншими критеріями виявлення змін поведінки об’єктів. Визначено, що в складних випадках критерій показує не гірший, а часто – кращий результат як за кількістю помилок першого роду та другого роду, так і за часом, що витрачається на прийняття рішення. Введення в практику моніторингу стану інформаційно-комп’ютерних мереж вказаного критерію дозволить підвищити рівень їх захищеності від DoS-атак різного типу, від зовнішнього проникнення, а також від інших причин втрати працездатності.

Ключові слова: аномальна поведінка інформаційно-комп'ютерної мережі, зміна моделі поведінки об’єкту, часовий ряд, автокореляція, критерій зміни моделі часового ряду.

Постановка задачі. Сучасні інформаційно-комп’ютерні мережі (ІКМ) можна розглядати як складні системи, про стан яких судять за значеннями ряду параметрів, що вимірюються. В процесі експлуатації ІКМ в завдання обслуговуючого персоналу входить постійний моніторинг її працездатності. При виникненні різного виду відмов обладнання та програмного забезпечення або інших позаштатних ситуацій ІКМ може істотно знижувати рівень працездатності (аж до повної відмови). Тому здатність вчасно зафіксувати факт виникнення аномальної поведінки ІКМ, визначити місце та/або її причину з метою подальшого відновлення працездатності є одним з найважливіших завдань як самого обслуговуючого персоналу, так і спеціалізованих автоматизованих засобів, покликаних спростити його роботу [1].

Відомо, що однією з основних причин втрати працездатності ІКМ сьогодні є атаки на них з боку зловмисників. Встановлено, що для трьох основних видів атак на ІКМ, які прийнято виділяти в даний час – розвідка, експлойт і відмова в обслуговуванні - зміна поведінки може бути виявлена саме через аналіз зміни параметрів, які описують її поведінку [1] - [4]. Оскільки спостереження за станом ІКМ звичайно проводиться впродовж визначеного проміжку часу, виміри значень параметрів, за якими ведеться спостереження представляються у формі моделей часових рядів. Відповідно, задача виявлення вторгнення та/або втрати працездатності

ІКМ зводиться до задачі виявлення зміни моделі часового ряду та формально описується наступним чином:

Маємо: запис сигналу Y у вигляді послідовних значень часового ряду {y_1,y_2,.....y_N}.

Необхідно: встановити, яка з двох визначених нижче гіпотез H_0 або H_1 є істинною.

H_0: всі значення y_1,y_2,.....y_N описано єдиною моделлю поведінки $\mathfrak{M}_0$ об'єкта або системи;

H_1: існує момент часу r: 1<r<N, такий що значення y_1,y_2,...y_{r-1} описуються моделлю $\mathfrak{M}_1$, а значення y_r.....y_N – моделлю $\mathfrak{M}_2$;

Якщо гіпотеза H_1 виявляється істинною, то додатково може ставитися задача оцінки моменту часу *r*, в який відбувається зміна моделі поведінки об'єкту. Розвиток даної схеми припускає, що потрібно визначити власне моделі $\mathfrak{M}_1$ та $\mathfrak{M}_2$.

В рамках даної роботи досліджується один з критеріїв прийняття рішення щодо зміни поведінки об'єкту дослідження, а саме такий, що базується на використання коефіцієнту автокореляції часового ряду першого порядку. Проводиться вивчення його можливостей, ефективності та обмежень, а також порівняння з іншими критеріями аналогічного призначення.

Огляд деяких можливих форм змін поведінки об'єктів [5] - [9]. Відомо, що різноманітність як властивостей самих моделей часових рядів, що представляють інформацію в галузі інформаційної безпеки, так і типів можливих змін в їх поведінці не дають змогу використати єдиний, універсальний метод виявлення цих змін. Наприклад, при описі об'єктів, що пов'язані з забезпеченням інформаційної безпеки, часто виникають в тому числі розподілення значень часових рядів, які суттєво відмінні від нормального закону. Зокрема, розмір кадрів, що передаються в мережі, сумарний обсяг інформації, що передається в мережі за одиницю часу здебільшого підпорядковуються логнормальному закону розподілення [5], моменти запитів на передачу кадрів в ІКМ є випадковим процесом, що підпорядковуються експоненціальному закону розподілення [6], а відношення обсягів вхідного та вихідного трафіку відповідають гамма-розподіленню [7].

З іншого боку заздалегідь невідомі й типи зміни моделі. Наприклад, показана на рис. 1. зміна моделі поведінки часового ряду відображає зміну положення або математичного очікування моделі ряду, що описує поведінку параметру. Саме такий випадок добре вивчений теоретично та практично, а його виявлення не представляє великих труднощів. На рис. 2. відображений випадок, який формально описується так само, як і попередній. Однак на відміну від нього зміна величини математичного очікування моделі значно менша (в наведеному випадку вона становить менше двох величин дисперсії модельного ряду). Це суттєво ускладнює виявлення (як алгоритмічне, так і таке, що виконується людиною-оператором) змін в поведінці об'єкту спостереження.

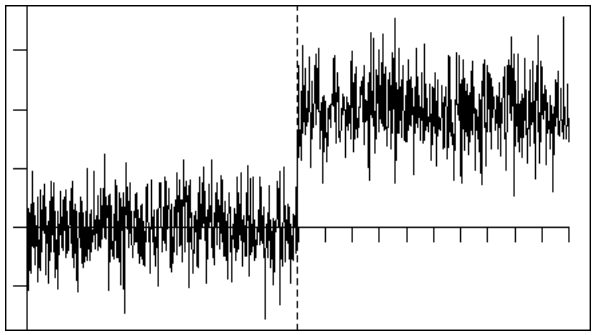


Рисунок 1 – Миттєва зміна параметру положення

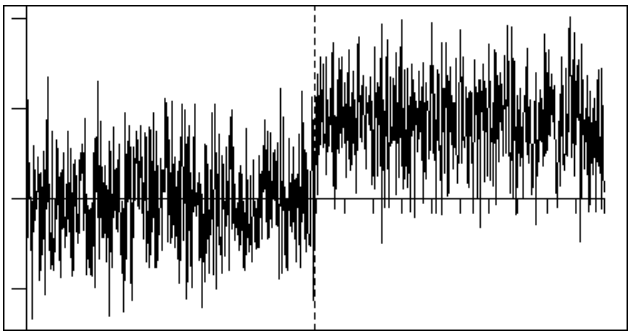


Рисунок 2 – Незначна зміна параметру положення

На рис.3 показано зміну параметру масштабу або дисперсії моделі. Цей тип змін характерний для багатьох аварійних ситуацій, зокрема в галузі технічної діагностики та відповідає випадку, коли об'єкт втрачає стійкість. Хоча параметр математичного очікування моделі не змінюється, ситуація часто являється провісником подальшого повного виходу об'єкту з ладу. На рис. 4. показаний більш складний для аналізу випадок, коли одночасно змінюються і параметр положення, і параметр масштабу моделі.

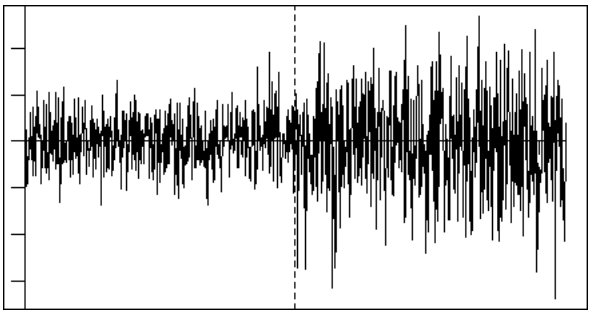


Рисунок 3 – Зміна параметру масштабу

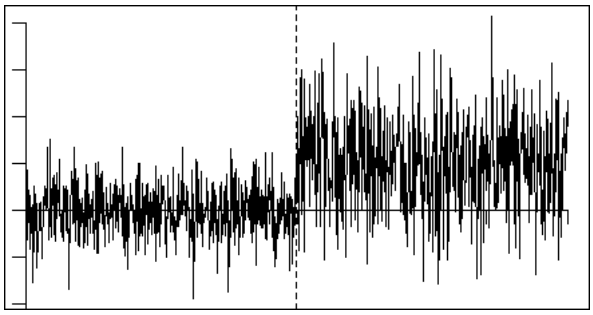


Рисунок 4 – Одночасна зміна параметрів положення та масштабу

Усі перераховані моделі відповідають ситуаціям, коли об’єкт переходить з одного стабільного стану в інший, тобто вказані параметри хоч і змінюються, але після моменту зміни впродовж досить тривалого часу зберігають нові параметри. На відміну від цього випадок, що відображений на рис.5, відображає ситуацію, коли після зміни стану модель перестає мати фіксоване значення параметру положення, тобто в моделі поведінки з’являється трендова складова. Виявити такі зміни набагато складніше ніж попередні. Однак, для практичних потреб уміння їх виявити – є надважливим, бо дає змогу вчасно відреагувати на появу тренду та не допустити виходу значень параметру за наперед задану граничну межу (наприклад, виявити DDoS атаку або діяльність шкідливого програмного забезпечення на ранніх стадіях їх активності). На мал.6 наведений приклад ускладнення попереднього випадку, який характеризується тим, що об’єкт взагалі не відзначається стабільними значеннями параметрів, тобто тренд має місце як до, так і після вказаної зміни, однак швидкість трендового процесу в різні проміжки часу різна. Наприклад, якщо параметр описує кількість відвідувачів інтернет-ресурсу, яка поступово збільшується зі збільшенням популярності ресурсу, то зміна тренду, якщо вона не співвідноситься з публікацією якихось надпопулярних матеріалів, редизайном сайту або застосуванням інших засобів привертання відвідувачів, може бути сигналом неконтрольованого підвищення уваги до сайту, можливо одного з різновидів хакерської атаки. Інший приклад зазначеного характеру поведінки об’єкту – перехід від поступового природньо-деградаційного характеру функціювання технічного об’єкту до етапу його катастрофічного руйнування. На рис.7 показаний випадок, коли основні параметри моделей – математичне очікування та дисперсія – лишаються сталими як до, так і після моменту зміни, але змінюється закон розподілення значень рядів.

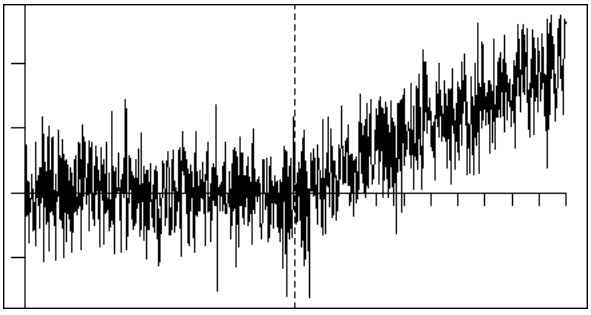


Рисунок 5 – Поява тренду

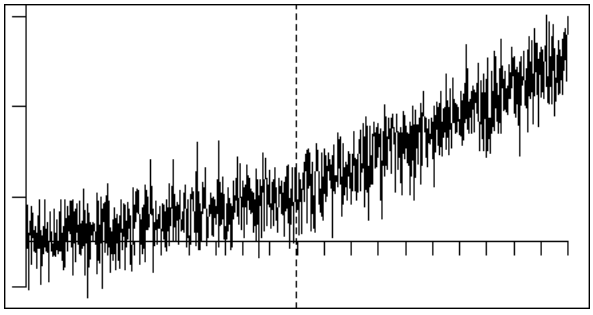


Рисунок 6 – Прискорення тренду

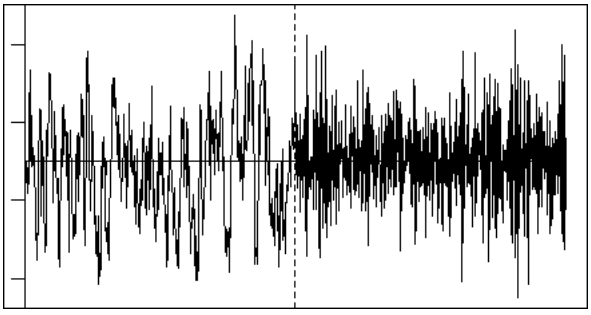


Рисунок 7 – Зміна закону розподілення

На сьогоднішній день лише найпростіші випадки зміни моделей піддаються формально-математичному опису та вивченню. В тих випадках, коли складність моделей, що виникають при вивченні практичних задач не дає змоги застосувати формалізацію, використовуються підходи, пов'язані з аналізом ситуацій за допомогою сучасних методів математичного моделювання та аналізу даних.

Таким чином, в арсеналі дослідження доцільно мати багато різнопланових методів та алгоритмів, які в сукупності дали б змогу виявити якомога більше різновидів мережних атак. Один з напрямів вирішення зазначеної проблеми [8] полягає в одночасному використанні ряду методів та прийняття рішення про зміну моделі поведінки за результатом їх спільного аналізу. Однак, це не знижує важливості вивчення інших, менш поширених на сьогодні критеріїв виявлення змін моделі поведінки та впровадження їх в практику моніторингу стану ІКМ.

Деякі існуючі методи виявлення змін моделей поведінки об'єктів [9], [10]. Серед параметричних методів виявлення зміни моделі поведінки ряду найбільш широковживаними є методи, які припускають, що зміна може бути визначена на основі аналізу таких параметрів моделі $\mathfrak{M}_0$, як математичне очікування та дисперсія. При цьому узагальнений алгоритм виявлення зміни моделі полягає в наступному:

1. Розбити ряд *Y* на два частини $Y^{(1)} = \{y_1, y_2, \ldots y_{n_1}\}$ та $Y^{(2)} = \{y_{\tau+1}, y_{\tau+2}, \ldots y_{\tau+n_2}\}$.
2. Сформулювати H_0 та H_1 гіпотези.
3. За допомогою відповідних критеріїв визначити, справдилися чи ні сформульовані вище гіпотези.
4. Зробити висновки про те чи змінилася модель поведінки часового ряду та надати цю інформацію в розпорядження дослідника для відповідної інтерпретації.
   (В подальшому, виходячи з можливостей практичного застосування описаних методів та алгоритмів, будемо вважати, що $n_1 = n_2 = n$).

Наприклад, якщо контролюються зміни математичного очікування моделей до та після змін в її поведінці, то гіпотези можна представити як

$$H_0 : E(Y^{(1)}) = E(Y^{(2)}),$$
$$H_1 : \left| E(Y^{(1)}) - E(Y^{(2)}) \right| > 0,$$

де $E(Y^{(1)})$ та $E(Y^{(2)})$ – математичне очікування моделей $\mathfrak{M}_1$ та $\mathfrak{M}_2$ відповідно.

У випадку, коли контролюються зміни дисперсій моделей, гіпотези представляються як

$$H_0 : D(Y^{(1)}) = D(Y^{(2)})$$
$$H_1 : D(Y^{(1)}) \neq D(Y^{(2)})$$

де $D(Y^{(1)})$ та $D(Y^{(2)})$ – дисперсія моделей $\mathfrak{M}_1$ та $\mathfrak{M}_2$ відповідно.

Пункт 1 зазначеного алгоритму не визначає явно, як саме мають бути сформовані ряди $Y^{(1)}$ та $Y^{(2)}$. Оскільки початок вимірювання значень загального часового ряду *Y* починається у момент отримання першого значення ряду $Y^{(1)}$, а рішення про наявність чи відсутність відмінності в поведінці рядів приймається після закінчення вимірювання останнього значення ряду $Y^{(2)}$, бажано, щоб між цими моментами проходило якомога менше часу (час затримки спостереження був мінімальний).

Однак, кількість значень рядів *n* обмежена знизу тим відомим фактом, що точність виявлення будь-якого статистичного параметру при малій кількості спостереженнь може бути недопустимо малою. Оцінка нижнього допустимого значення *n* має бути проведена для виявлення статистичної потужності критерію прийняття/відхилення гіпотези.

Інший шлях зменшення часу затримки спостереження – сумістити ряди $Y^{(1)}$ та $Y^{(2)}$ між собою. Суміщення значень рядів $Y^{(1)}$ та $Y^{(2)}$, що є підмножинами сукупного ряду $Y = Y^{(1)} \bigcup Y^{(2)}$ означає, що існує індекс *l* елементів ряду $Y^{(1)}$ такий, що $y^{(1)}_{l+k} = y^{(1)}_{l+k}$, $k = 0,.., n-l$, тобто $Y^{(1)} \bigcap Y^{(2)} \neq \varnothing$. При $l=1$ отримуємо ряди, що повністю співпадають, а отже говорити про зміну моделі не виявляється можливим. При $l>n$ маємо два ряди, що не перетинаються, при цьому

величина вікна спостереження (а отже і час затримки на роботу алгоритму виявлення зміни моделі) дорівнює *2n*. Вибір оптимального з точки зору мінімізації часу затримки та максимізації статистичної потужності алгоритму значення *l* є окремою задачею.

Використання коефіцієнту автокореляції для виявлення змін моделі поведінки. Окрім зазначених вище параметрів моделей часових рядів, доцільно звернути увагу на ще один їх параметр, а саме на коефіцієнт автокореляції. На відміну від параметру положення (математичного очікування) та параметру масштабу (дисперсії), унікальність коефіцієнту автокореляції проявляється в тому, що він здатен виявляти залежність між значеннями ряду в часі через виявлення залежності послідовних значень ряду між собою. Отже, доцільно вивчити можливість застосування вказаного параметру при вирішенні задачі виявлення зміни моделі поведінки часового ряду.

Коефіцієнтом автокореляції першого порядку називається статистика, що визначається наступним чином:

$$r = \frac{\sum_{t=2}^{n}\left(y_t - \overline{y}_1\right)\cdot\left(y_{t-1} - \overline{y}_2\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \overline{y}_1\right)^2 \cdot \sum_{t=2}^{n}\left(y_{t-1} - \overline{y}_2\right)^2}}, \quad (1)$$

де $\overline{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1}; \overline{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1};$

Особливістю вибіркового коефіцієнту кореляції як імовірнісної характеристики часового ряду є досить складний характер її функції розподілення. Тому на різних діапазонах значень використовується різна її апроксимація. Так при *r*, близьких до 0, тобто при перевірці гіпотези $H_0 : r = 0$ проти гіпотези $H_1 : r \neq 0$, що семантично еквівалентно перевірці гіпотези про відсутність кореляції між послідовними значеннями ряду *Y*, статистика

$$R = r\frac{\sqrt{n-2}}{1-r^2} \quad (2)$$

досить точно апроксимується законом розподілення Стьюдента з *df=(n-2)* ступенями свободи. При збільшенні інтенсивності кореляційного зв'язку, тобто при значеннях *r*, що наближені до ± 1, розподілення вибіркового коефіцієнту кореляції стає все більш асиметричним і його збіжності до нормального відповідно зменшується. Щоб перешкодити цьому можна скористуватися нормуючим перетворенням Фішера [9], а саме замість *r* розглядати величну z, що розраховується за наступною формулою:

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} \quad (3)$$

При $n \to \infty$ математичне очікування та дисперсія випадкової величини, окремі реалізації якої розраховуються за наведеною формулою, описуються як

$$E(z) = \frac{1}{2}\ln\frac{1+r}{1-r} + \frac{r}{2(n-1)} \quad (4)$$

та

$$D(z) = \frac{1}{n-3} \quad (5)$$

Головною особливістю випадкової величини *z* є те, що її дисперсія залежить виключно від кількості проведених спостережень *n* та не залежить від значення *r*. В описаному вигляді коефіцієнт автокореляції та його статистика *z* досить часто застосовується для виявлення наявності (або відсутності) трендової складової в часових рядах.

Розглянемо можливість застосування цього критерію для вирішення задачі виявлення зміни в поведінці моделі часового ряду на основі порівняння значень його параметрів. Позначимо $r_{(1)}$, $r_{(2)}$ – вибіркові коефіцієнти кореляції рядів $Y^{(1)}$ та $Y^{(2)}$, що розглядаються як такі, що отримані

з генеральних сукупностей, моделі поведінки яких мають коефіцієнти кореляції $\rho_{(1)}$ та $\rho_{(2)}$ відповідно. Тоді нульова гіпотеза може бути сформульована як

$$H_0 : \rho_{(1)} = \rho_{(2)},$$

а відповідна альтернативна гіпотеза

$$H_1 : \rho_{(1)} \neq \rho_{(2)}.$$

Таким чином, задача зведена до виявлення, чи є різниця $ZR = \left| r_{(1)} - r_{(2)} \right|$ достатньо великою, щоб відхилити гіпотезу H_0.

Виходячи з співвідношень (3) - (5) отримуємо, що значення випадкової величини, яка визначаються формулою

$$ZR = \frac{1}{2}\left| \ln\frac{1+r_{(1)}}{1-r_{(2)}} - \ln\frac{1+r_{(2)}}{1-r_{(2)}} \right| \Big/ \sqrt{\frac{2}{n-3}} \qquad (6)$$

мають бути розподілені відповідо до нормального закону розподілення *N(0,1)*.

Статистичне моделювання для виявлення можливостей критерію на основі коефіцієнту автокореляції часового ряду. Перше питання, яке необхідно вирішити для вивчення можливості використання статистики (6) – це визначити межі її використання з точки зору обмеження знизу на величину вікна спостереження *2*n*, тобто таких їх значень, при яких перестають спостерігатися суттєві відхилення розподілення випадкової величини *ZR* від стандартного нормального закону розподілення. Крім того, необхідно визначити потужність статистики *ZR* при виявленні наявності/відсутності автокореляції в залежності від значення коефіцієнту автокореляції генеральної сукупності ρ.

Для аналізу буда проведена серія з 5000 статистичних експериментів, у кожному з яких генерувалися розподілення з заданим значенням ρ з множини 0.01,0.1,0.2,0.5,0.8,0.9}, вибирався величина вікна розподілення *2*n* з множини {10,20,30,60,90,120}, та визначалося значення *p-value*, отримане по кожному з наступних критеріїв виявлення відповідності значень вибірок *ZR* стандартному нормальному закону розподілення: Шапиро-Уилка, хи-квадрат Пірсона, та Крамера-фон Мизеса [10]. (Цей та всі наступні описані статистичні експерименти провадилися за допомогою пакету аналізу *R*, v.3.2.2 з застосуванням середовища RStudio, v. 0.99.486).

Використавши методику, яка описана в роботі [11], визначимо усереднене значення досягнутих рівнів значущості зазначених критеріїв. Результати експериментів наведені в табл.1.

Таблиця 1 – Середні досягнуті значення *p-value*

Значення	**Величина вікна спостереження**						
ρ	**10**	**16**	**20**	**30**	**60**	**90**	**120**
0	3.00E-10	0.0439369	0.3190211	0.3955553	0.4960906	0.4479223	0.4440315
0.01	2.50E-10	0.0663711	0.3225771	0.4067907	0.4890609	0.5118037	0.4832334
0.1	2.95E-10	0.0618481	0.3385225	0.4539149	0.4753561	0.5187863	0.4848777
0.2	3.00E-10	0.0507316	0.3394343	0.4422184	0.5410921	0.4740081	0.5424239
0.5	3.48E-09	0.0487950	0.3413038	0.4487951	0.4875496	0.5038067	0.4714025
0.8	4.31E-05	0.0906523	0.3571873	0.4196046	0.4417202	0.4778238	0.4893302
0.9	2.39E-03	0.1272577	0.3908079	0.4087724	0.3929194	0.4469495	0.4694467

З табл.1 видно, що вже при величині вікна спостереження що дорівнює 20 стабільно забезпечується рівень значущості, який перевищує помилку першого роду $\alpha = 0{,}05$. При цьому помітно, що для вікон спостереження ≥20 середнє досягнуте значення *p-value* зростає з збільшенням ρ, досягаючи максимуму на значеннях ρ близьких до 0.5.

При визначенні придатності критерію до розв'язання практичних задач важливим показником є кількість помилок першого та другого роду, що він допускає. Для цього була

проведена серія з 5000 експериментів, для кожного з якого вибирались можливі пари значень коефіцієнтів $\rho^{(1)}$ та $\rho^{(2)}$ (з множини {0.1,…0.9}, при умові $\rho^{(1)} \geq \rho^{(2)}$) та величини вікна спостереження *2*n* з множини {20,30,60,90}. Відносна кількість помилкових спрацювань для деяких значень *2*n* та різниці $\Delta\rho=\rho^{(2)}-\rho^{(1)}$ наведена в табл. 2-6.

Таблиця 2 – Відносна кількість помилкових спрацювань при $\Delta\rho=0.1$

Вікно спостереження	**Зміна коефіцієнту *ρ* моделей (від значення –до значення).**								
	0.0-0.1	**0.1-0.2**	**0.2-0.3**	**0.3-0.4**	**0.4-0.5**	**0.5-0.6**	**0.6-0.7**	**0.7-0.8**	**0.8-0.9**
20	00.071	00.075	00.075	00.075	00.086	00.096	00.113	00.121	00.148
30	00.080	00.086	00.089	0.0103	00.109	00.132	00.142	00.184	00.206
60	00.110	00.117	00.122	00.125	00.150	00.196	00.223	00.277	00.332
90	00.125	00.123	00.149	00.156	00.175	00.209	00.265	00.355	00.423

Таблиця 3 – Відносна кількість помилкових спрацювань при $\Delta\rho=0.2$

Вікно спостереження	**Зміна коефіцієнту ρ моделей (від значення –до значення).**							
	0.0-0.2	**0.1-0.3**	**0.2-0.4**	**0.3-0.5**	**0.4-0.6**	**0.5-0.7**	**0.6-0.8**	**0.7-0.9**
20	00.085	00.084	00.085	00.097	00.103	00.126	00.134	00.166
30	00.106	00.110	00.126	00.136	00.155	00.173	00.204	00.232
60	00.176	00.184	00.196	00.233	00.255	00.301	00.350	00.419
90	00.214	00.240	00.260	00.277	00.324	00.385	00.485	00.560

Таблиця 4 – Відносна кількість помилкових спрацювань при $\Delta\rho=0.5$

Вікно спостереження	**Зміна коефіцієнту ρ моделей (від-до).**				
	0.0-0.5	**0.1-0.6**	**0.2-0.7**	**0.3-0.8**	**0.4-0.9**
20	0.170	0.180	0.180	0.199	0.219
30	0.274	0.294	0.331	0.337	0.393
60	0.560	0.606	0.631	0.694	0.738
90	0.745	0.777	0.826	0.868	0.908

Таблиця 5 – Відносна кількість помилкових спрацювань при $\Delta\rho=0.8$

Вікно спостереження	**Зміна коефіцієнту *ρ* моделей (від-до).**	
	0-0.8	**0-0.9**
20	0.316	0.326
30	0.568	0.563
60	0.912	0.925
90	0.983	0.988

Таблиця 6 – Відносна кількість помилкових спрацювань при $\Delta\rho=0.9$

Вікно спостереження	**Зміна коефіцієнту ρ моделей (від-до).**
	0-0.9
20	0.387
30	0.650
60	0.961
90	0.996

У табл.7. наведені результати визначення потужності критерію, тобто кількості випадків правильного розпізнавання зміни моделі поведінки часового ряду в залежності від значень коефіцієнтів автокореляції *ρ(1)* та *ρ(2)* моделей $\mathfrak{M}1$ та $\mathfrak{M}2$.

Очікувано потужність критерію збільшується зі збільшенням різниці між $\rho^{(1)}$ та $\rho^{(2)}$ та від розміру вікна спостереження *2*n*. Оскільки в реальних задачах ця різниця нам апріорі невідома, наведена також середня відносна кількість отриманих вірних рішень. Аналіз наведених в таблиці значень показує, що використання розміру вікна, меншого ніж 2*n=60 спостережень при невідомих законах розподілення рядів навряд чи виглядає доцільним.

Таблиця 7 – Відносна кількість правильно розпізнаних змін моделі

Вікно спостереження	Значення $\rho^{(2)}$-$\rho^{(1)}$									Середня відносна кількість вірних рішень
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
20	0.0955	0.1100	0.1312	0.1503	0.1898	0.2288	0.2809	0.3208	0.3874	0.1621
30	0.1256	0.1553	0.1987	0.2491	0.3260	0.4051	0.4876	0.5655	0.6504	0.2612
60	0.1835	0.2643	0.3766	0.5001	0.6460	0.7684	0.8583	0.9189	0.9610	0.4684
90	0.2200	0.3430	0.5117	0.6765	0.8247	0.9187	0.9669	0.9855	0.9962	0.5785

Моделювання виявлення зміни моделі поведінки об'єкту на основі зміни коефіцієнту автокореляції. Проведений аналіз дозволяє перейти до безпосереднього моделювання виявлення зміни моделі часового ряду на основі аналізу коефіцієнту автокореляції, та його порівняння з іншими критеріями, зокрема, на основі аналізу математичного очікування та аналізу дисперсії. Для цього була проведена серія статистичних експериментів, для чого кожен приклад зміни моделі поведінки об'єкту з наведених на рис.1-7 моделювався 25000 разів, в кожному генерувалися ряд Y обсягом 300 точок: $Y = \{y_1, y_2, .., y_{300}\}$. Алгоритм аналізування, що використовувався на кожному з 25000 циклів експерименту полягав в наступному:

1. Вибрати вікно спостереження *win*.
2. Для всіх значень *i* таких, що належать діапазону від *win* до 300
   - сформувати ряд $Y^{(1)}$ з послідовних значень ряду Y: $\left\{ y_{(i-win+1)}, ... y_{(i-\frac{win}{2})} \right\}$;
   - сформувати ряд $Y^{(2)}$ з послідовних значень ряду Y: $\left\{ y_{(i-\frac{win}{2}+1)}, ..., y_{(i)} \right\}$;
   - застосувати до $Y^{(1)}$ та $Y^{(2)}$ *t*-тест Стьюдента для виявлення зсуву параметру положення;
   - застосувати до $Y^{(1)}$ та $Y^{(2)}$ *F*-тест Фішера для виявлення зміни параметру масштабу;
   - застосувати до $Y^{(1)}$ та $Y^{(2)}$ критерій на основі аналізу коефіцієнту автокореляції.
3. Для кожного з використаних тестів підрахувати кількість помилкових спрацювань та номер першого спрацювання тесту після моменту зміни моделі часового ряду.
4. У разі, якщо не всі величини вікон спостереження були використані, вибрати наступне значення *win* та перейти до п.2.

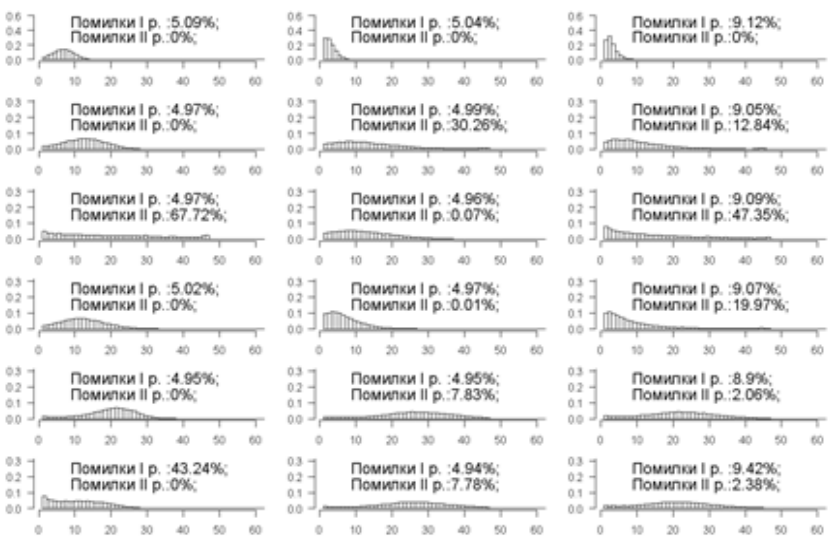
Таким чином, алгоритм застосовує послідовне просування вікна спостереження вздовж послідовності значень ряду Y та розрахунок для кожного з критеріїв кількості помилок першого роду (кількість помилкових сигналів), другого роду (кількість випадків, коли зміна так і не була зафіксована) та затримки (у вигляді кількості відрахувань значень ряду) що потребував критерій для виявлення зміни моделі після того, як така зміна дійсно мала місце. Для визначення факту спрацювання приймався факт отримання значення *p-value* меншого за 0.05 (5% поріг значущості). Відповідно до висновків, що були зроблені раніше, використовувалися вікна спостереження 60 та 90.

Результати моделювання відображені в табл.8. та 9. Аналізуючи табл.8. бачимо, що в найпростіших випадках – миттєва зміна параметру положення або параметру масштабу – очікувано найкращі результати за показником мінімізації помилок першого та другого роду показують відповідно *t*-тест Стьюдента та *F*-тест Фішера. В більш складних випадках перевага вказаних критеріїв суттєво зменшується, а в деяких випадках показники критерію з використанням коефіцієнту автокореляції перевищують показники базових тестів. Таким чином, точно визначити, який з методів виявлення зміни моделі поведінки краще застосовувати можна тільки в випадку, коли заздалегідь відомий можливий тип змін. В реальних ситуаціях така інформації у особи або системи, що провадить моніторинг стану об'єкту, відсутня. Деяке уявлення про узагальнені результати моделювання можна отримати, якщо усереднити значення показників тестів для суміші всіх типів зміни поведінки моделі, що розглядалися, або взявши до уваги лише найскладніші з точки зору розпізнавання випадки.

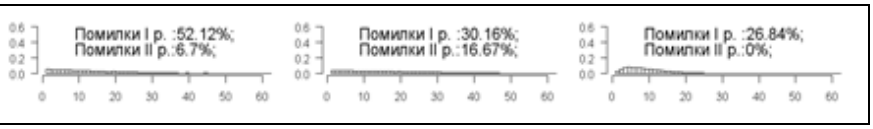
Таблиця 8 – Результати моделювання алгоритмів виявлення зміни моделі поведінки часових рядів

№ з/с	Тип зміни моделі поведінки об'єкту спостереження	Розмір вікна спостереження = 60						Розмір вікна спостереження = 90					
		Алгоритми визначення зміни моделі, що використовувалися											
		t-тест Стьюдента		F-тест Фішера		Тест на основі аналізу коефіцієнту автокореляції		t-тест Стьюдента		F-тест Фішера		Тест на основі аналізу коефіцієнту автокореляції	
		Відносна кількість виявлених помилок											
		I роду	II роду	I роду	II роду	I роду	II роду	I роду	II роду	I роду	II роду	I роду	II роду
1.	Миттєва зміна параметру положення	0.0504	0.0000	0.0507	0.0000	0.0864	0.0006	0.0508	0.0000	0.0504	0.0000	0.0912	0.0000
2.	Незначна зміна параметру положення	0.0497	0.0000	0.0502	0.4148	0.0862	0.2258	0.0497	0.0000	0.0499	0.3026	0.0905	0.1284
3.	Зміна параметру масштабу	0.0495	0.6870	0.0497	0.0118	0.0859	0.5112	0.0497	0.6772	0.0496	0.0007	0.0909	0.4735
4.	Одночасна зміна параметрів положення та масштабу	0.0498	0.0004	0.0497	0.0028	0.0859	0.2890	0.0502	0.0000	0.0497	0.0001	0.0907	0.1997
5.	Поява тренду.	0.0495	0.0003	0.0500	0.4767	0.0852	0.2480	0.0495	0.0000	0.0495	0.0783	0.0890	0.0206
6.	Прискорення тренду.	0.1636	0.0000	0.0496	0.4791	0.0872	0.2494	0.4324	0.0000	0.0495	0.0778	0.0942	0.0238
7.	Зміна закону розподілення.	0.5187	0.0841	0.2720	0.2150	0.2413	0.0000	0.5212	0.0670	0.3016	0.1667	0.2684	0.0000
Середне значення попомилки на повній суміші прикладів		0.1330	0.1103	0.0817	0.2286	0.1083	0.2177	0.1719	0.1063	0.0857	0.0894	0.1164	0.1209
Середне значення попомилки на суміші прикладів 2,4,6,7		0.1954	0.0211	0.1054	0.2779	0.1252	0.1911	0.2634	0.0168	0.1127	0.1368	0.1360	0.0880

Таблиця 9 – Результати моделювання. Помилки I та II роду та затримка в прийнятті рішення при розмірі вікна спостереження = 90

№ з/с	Тип зміни моделі поведінки об'єкту спостереження	t-тест Стьюдента	F-тест Фішера	Тест на основі аналізу коефіцієнту автокореляції
1.	Миттєва зміна параметру положення	
2.	Незначна зміна параметру положення			
3.	Зміна параметру масштабу			
4.	Одночасна зміна параметрів положення та масштабу			
5.	Поява тренду			
6.	Прискорення тренду			

Продовженя таблиці 9

7.	**Зміна закону розподілення**	

Окрім помилок першого та другого роду, що допускаються тестами, важливе значення в реальних процесах має такий показник як час затримки між моментом фактичної зміни поведінки об’єкту спостереження та часу виявлення зміни (наприклад, між моментом початку DDoS атаки та моментом, коли обслуговуючий персонал отримав відповідний сигнал). У табл.9. наведені результати моделювання розподілу часу затримки в залежності від критерію, що використовувався та типу зміни моделі при розмірі вікна спостереження, що дорівнює 90. Там же наведені значення помилок I та II роду. На гістограмах розподілення видно, що в деяких випадках тест на основі аналізу коефіцієнту автокореляції хоча і показує значення помилок дещо гірші за інші тести, але розподілення часу суттєво скошене в сторону малих значень. Такий скіс означає, що час, який потребувався для фіксації факту зміни моделі, був менший. Зазначений скіс спостерігається навіть в “зручних” для аналізу випадках, таких як 1 та 3 для t-тесту Стьюдента та F-тесту Фішера відповідно. Оскільки в реальних ситуаціях передбачити, як саме буде змінюватися модель поведінки об’єкту, як правило неможливо, розглянутий критерій на основі аналізу коефіцієнту автокореляції першого порядку часового ряду може бути включений до набору тестів сумісного використання для підвищення загальної ефективності системи розпізнавання позаштатних ситуацій.

Висновки. З результатів проведених статистичних експериментів випливає, що критерій на основі аналізу коефіцієнту автокореляції першого порядку володіє достатньою потужністю та ефективністю для використання при визначенні факту зміни моделі поведінки об’єкту. При цьому, хоча в найпростіших випадках він показує результати, що поступаються іншим критеріям, в деяких складних випадках він перевищує їх можливості. Крім того, критерій здатен забезпечити значення часу затримки в прийнятті рішення менший в порівняння з конкурентними критеріями. Отже, можна вважати доцільним при побудові систем розпізнавання позаштатних ситуацій в ІКМ, в тому числі систем виявлення вторгнень, включати цей критерій, а також реалізовувати алгоритми прийняття рішень на основі спільного аналізу сукупності результатів використання групи критеріїв.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] О.И.Шелухин, Д.Ж.Сакалема, и А.С.Филинова, *Обнаружение вторжений в компьютерные сети (сетевые аномалии)*. Москва, РФ: Горячая линия-Телеком,2013.

[2] N.Adams, and N.Heard, *Data analysis for network cyber-security*. Singapor: Imperial College Press, 2014.

[3] M.Collins, *Network Security Through Data Analysis*. Sebastopol, CA, USA: O’Reilly Media Inc., 2014.

[4] H.Wang, D. Zhang , and K.G.Shin, "Change-Point Monitoring for Detection of DoS Attacks", *IEEE Transactions on Dependable and Secure Computing*, vol. 1, is. 4., pp.193 - 208, 2004.

[5] В.В.Петров, Статистический анализ сетевого трафика. [Электронный ресурс]. Доступно: http:// www.pi.314159.ru/petroff2.pdf, Дата звернення: Квіт.12, 2017.

[6] В.Л.Тамп, Н.В.Тамп, и А.А.Кузьмин, "Иммитационная модель потоков запросов на передачу кадров в информационно-вычислительной сети", *Вестник Череповецкого ГУ,* №8, с.32-35, 2015.

[7] Р.Р.Фактиева, “Разработка метрик для обнаружения атак на основе анализа сетевого трафика”, *Вестник Бурятского ГУ,* №3, с.81-86, 2013.

[8] В.С.Ловягин, “Статистический мониторинг вирусных атак на основе параметрических критериев”, *Вісник Севастопільського НТУ: зб. наук. пр., Серія: Інформатика, електроніка, зв'язок,* Вип. 114, с.31-35, 2011.

[9] С.А.Айвазян, И.С.Енюков, и Л.Д.Мешалкин, *Прикладная статистика: Исследование зависимостей,* Под ред. С.А.Айвазяна, Москва, СССР: Финансы и статистика, 1985.
[10] А.И.Кобзарь, *Прикладная математическая статистика. Для инженеров и научных работников,* Москва,РФ: ФИЗМАТЛИТ, 2006.
[11] В.М.Волкова, "Исследование распределения статистик критерия обнаружения сдвига средних Кохрана", *Вестник Томского ГУ, Управление, вычислительная техника и информатика,* №1(26),-с.31-38, 2014.

Стаття надійшла до редакції 24 березня 2017 року.

REFERENCE

[1] O.I.Sheluhin, D.J. Sakalama, A.S.,Filinova, *Intrusion Detection in computer networks (network anomalies).* Moscow, Russia, hotline-Telecom,2013.
[2] N.Adams, and N.Heard, *Data analysis for network cyber-security.* Singapor: Imperial College Press, 2014.
[3] M.Collins, *Network Security Through Data Analysis.* Sebastopol, CA, USA: O'Reilly Media Inc., 2014.
[4] H.Wang, D.Zhang, and K.G.Shin, "Change-Point Monitoring for Detection of DoS Attacks", *IEEE Transactions on Dependable and Secure Computing,* vol. 1, is. 4., pp.193 - 208, 2004.
[5] V.V.Petrov, Statistical analysis of network traffic. [Online]. Available: http://www.pi.314159.ru/petroff2.pdf, Accessed on: Apr.12, 2017.
[6] V.L.Tamp, N.V.Tamp, and A.Kuzmin, "Simulation model of flows of requests for transfer of personnel in an information network", *Bulletin of Cherepovets SU,* No.8, pp.32-35, 2015.
[7] R.R.Factieva, "Development of metrics for detection of attacks based on network traffic analysis", *Bulletin of the Buryat SU,* No. 3, pp. 81-86, 2013.
[8] V.S. Lovyagin, "Statistical monitoring of virus attacks based on parametric criteria", *Sevastopol STU: Collection of scientific papers, Series: computer science, electronics, communications,* Vol. 114, pp.31-35, 2011.
[9] S.A.Aivazyan, I.S.Enyukov, and L.D.Meshalkin, *Applied statistics: Research of dependences,* Under the editorship of S. A. Ayvazian, Moscow, USSR: Finansy&Statistika, 1985.
[10] A.I.Kobzar, *Applied mathematical statistics. For engineers and scientists,* Moscow,Russia: FIZMATLIT, 2006.
[11] V.M.Volkova, "The investigation of statistic distributions of the Cochran test for the means shift detection", *Bulletin of the Tomsk SU, Management, Computing and Informatics,* №1(26), pp.31-38, 2014.

ДМИТРИЙ ШАРАДКИН

ИСПОЛЬЗОВАНИЕ КРИТЕРИЯ ВЫЯВЛЕНИЯ ИЗМЕНЕНИЯ ПОВЕДЕНИЯ ОБЪЕКТА НА ОСНОВЕ АНАЛИЗА КОЭФФИЦИЕНТА АВТОКОРРЕЛЯЦИИ В ЗАДАЧАХ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Рассматриваются методы выявления факта изменения поведения объектов, в частности современных информационно-компьютерных сетей, на основе анализа их моделей функционирования в виде временных рядов. Показано, что указанные объекты характеризуются большой внутренней сложностью, а также разнообразием статистических законов распределения их значений. Самым сложным для исследования и практического применения является наличие широкого спектра возможных форм и характеристик изменений поведения, вызванные непредсказуемостью как самих причинам, так и их возможного влияния на указанные объекты. Описанные ограничения не дают возможности выбрать единый универсальный критерий выявления изменений в поведении объектов исследования, который был бы способен реагировать на большинство изменений, и приводят к необходимости

совместного применения ансамбля соответствующих критериев. В работе подробно рассматривается один из таких критериев, а именно критерий, основанный на использование коэффициента автокорреляции временного ряда первого порядка. С помощью средств статистического моделирования проведен анализ основных характеристик критерия, изучены показатели его статистической мощности, эффективности, а также различные ограничения его использования. Проанализированы зависимости результатов использования критерия от параметров алгоритма обнаружения изменения. Проведено сравнение предложенного критерия с другими критериями выявления изменений поведения объектов. Определено, что в сложных случаях критерий показывает не хуже, а часто - лучший результат как по количеству ошибок первого рода и второго рода, так и по времени, которое затрачивается на принятие решения. Введение в практику мониторинга состояния информационно-компьютерных сетей указанного критерия позволит повысить уровень их защищенности от DoS-атак различного типа, от внешнего проникновения, а также от других причин потери работоспособности.

Ключевые слова: аномальное поведение информационно-компьютерной сети, изменение модели поведения объекта, временной ряд, автокорреляция, критерий изменения модели временного ряда.

DMYRTO SHARADKIN

CHANGE-POINT DETECTION TEST BASED ON THE ANALYSIS OF THE TIME SERIES' AUTOCORRELATION AND ITS APPLICATION FOR INFORMATION SECURITY

Methods for detection changes in the behavior of technical objects, in particular in modern information and computer networks, which are based on the analysis of time series has been investigated. It is shown that these objects are characterized by great internal complexity, as well as a variety of probability distribution of their values. A wide range of possible forms and characteristics of behavior changes caused by unpredictability of both the causes themselves and their possible impact on these objects makes research and practical application of change-point detection in this field extremely difficult. These limitations restrict every single method and require the combined aggregate application of the tests for change-point detection in models. The paper survey is one of such tests, which is based on the application of the first order autocorrelation coefficient of the time series. Statistical simulation of the process has been applied for analysis of the possibilities of the test, its power, efficiency and restrictions. Dependencies of the test's results on the various change-detection algorithm parameters are analyzed. An examining, analyzing and comparison of the test with similar ones for detection of changes in the behavior of objects has been executed. It was determined that in difficult cases the test shows not the worst, but often the best result in terms of the numbers of type I and type II errors, and of the time, which was spent for decision making. The utilization of this test for monitoring of the information and computer networks could increase the level of protection against various types of DoS attacks, intrusions, as well as from other causes of efficiency loss.

Keywords: computers network's anomalous behavior, change-points of the model, time series, autocorrelation, time series changes detection test.

Дмитро Михайлович Шарадкін, кандидат технічних наук, доцент, доцент кафедри кібербезпеки та застосування автоматизованих інформаційних систем та технологій, Інститут спеціального зв'язку та захисту інформації Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна.

E-mail: dmsh@ukr.net.

Дмитрий Михайлович Шарадкин, кандидат технических наук, доцент, доцент кафедры кибербезопасности и применения автоматизированных информационных систем и

технологий, Институт специальной связи и защиты информации Национального технического университета Украины “Киевский политехнический институт имени Игоря Сикорского”, Киев, Украина.

Dmyto Sharadkin, candidate of technical sciences, associate professor, associate professor at the cybersecurity and application of automated information systems and technologies academic department, Institute of special communication and information protection National technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

УДК 004(94+41)

ВОЛОДИМИР СОКОЛОВ

ЗАСТОСУВАННЯ ФУНКЦІОНАЛЬНОЇ ТА РЕЛЯЦІЙНОЇ МОДЕЛЕЙ В ОБ’ЄКТНО-ОРІЄНТОВАНОМУ ПРОГРАМУВАННІ

В роботі представлено результати досліджень та практичної апробації формальних методів опису об’єктно-орієнтованих програм, придатних для автоматичної генерації тексту програм на мові програмування. В якості формальних моделей обрано функціональну та реляційну моделі. Функціональна модель представляє програму як схему сполуки функціональних атомарних об’єктів, здатних до безпосередньої взаємодії шляхом утворення динамічних сполук та автоматичних обчислень, яка може представлятися у графічній формі. При цьому схема з’єднання об’єктів розглядається як схема розв’язку задачі. Показано, що формалізація саме схеми розв’язку задачі, а не всієї парадигми програмування, робить процес створення програми більш наближеним до практики. Визначені вимоги до атомарних об’єктів як таких, що складають елементну базу об’єктно-орієнтованої програми. Реляційна модель представляє об’єкт як віртуальне відношення, схема якого задається класом, який реалізує функціональну залежність неключових атрибутів від ключових шляхом їх обчислень, що дозволяє застосовувати реляційні операції для опису схеми розв’язку задачі. Реляційна модель дозволяє використовувати мову, подібну до структурованої мови запитів до баз даних, для опису схеми розв’язку задачі та її автоматичного виконання. Показано, що функціональна та реляційна моделі придатні для графічного представлення схеми розв’язку задачі і є достатньовиразними для безпосередньої генерації програм. Фактично, розроблені моделі дозволяють підняти процес створення об’єктно-орієнтованих програм на рівень вище, зосередитись на структурі програми, а не на її складових, і доповнити прогалину в існуючих методах представлення програм. В якості основи для практичної реалізації використана технологія програмування активних динамічних сполук об’єктів.

Ключові слова: об’єктно-орієнтоване програмування, формалізація програм, реляційна модель, функціональна модель, активні динамічні сполуки об’єктів.

Постановка проблеми. Об’єктно-орієнтоване програмування (ООП), незважаючи на його широке використання, має низку проблем, найсуттєвішими з яких є: наявність “крихкого” базового класу в успадкуванні, відсутність формальної теорії опису об’єктно-орієнтованих програм (ООПр) для їх верифікації, перевірки коректності, доведення вірності й оптимізації, а також відсутність засобів високо рівня побудови ООПр з об’єктів та уніфікованих механізмів їх взаємодії. Проблеми успадкування в першу чергу пов’язані з тим, що успадкований базовий клас потрапляє в простір імен похідного класу, що часто призводить до збігу імен, наявність в структурі пам’яті об’єкту похідного класу всього об’єкту базового класу та непередбачуваність впливу змін в базовому класі на всі класи-нащадки.