

Радченко К.О.

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

КОНЦЕПТУАЛЬНА МОДЕЛЬ ЗАБЕЗПЕЧЕННЯ ЕФЕКТИВНОСТІ ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ НА ВЕБСЕРВЕР

Тематика статті пов'язана з вирішенням актуальної науково-практичної задачі вдосконалення математичного забезпечення прогнозування навантаження на вебсервер. У результаті проведеного аналізу літературних джерел визначена перспективність вдосконалення шаблонів нормальної поведінки вебсерверів шляхом впровадження у їх математичне забезпечення сучасних методів частотно-часового аналізу сигналів на базі теорії вейвлет-перетворень.

Визначена недостатня висвітленість у доступній літературі питань, пов'язаних із обґрунтуванням адаптації типу базисного вейвлету до умов створення шаблону нормальної поведінки вебсерверу. Запропоновано виправити вказаний недолік шляхом створення методу визначення ефективного типу базисного вейвлету, призначеного для розробки шаблону нормальної поведінки. Для цього обґрунтовано низку положень і перелік критеріїв ефективності, які дозволяють забезпечити вибір ефективного типу базисного вейвлету відповідно до важливіших вимог задачі розробки вказаного шаблону.

Суть методу полягає у визначенні параметрів, які характеризують базисні вейвлети, та їх співвіднесенні з важливими умовами формування шаблону нормальної поведінки вебсерверу. За допомогою комп'ютерних експериментів показано, що метод дозволяє обрати тип базисного вейвлету, який забезпечує врахування у побудованому шаблоні нормальної поведінки складного багатоперіодичного характеру динаміки параметрів захисту вебсерверу. Точність такого шаблону знаходиться на рівні сучасних систем прогнозування навантаження вебсерверу. При цьому важливою перевагою методу є уникнення в процесі розробки вейвлет-моделі шаблону нормальної поведінки тривалих численних експериментів, спрямованих на визначення типу базисного вейвлету.

Перспективи подальших досліджень пов'язані з розвитком запропонованого методу в напрямі адаптації параметрів базисного вейвлету до умов поставленої задачі.

Ключові слова: вейвлет-модель, шаблон нормальної поведінки, вебсервер, захист інформації, базисний вейвлет.

Постановка проблеми. У зв'язку з інтенсивним зростанням мережевих комп'ютерних систем, впровадженням таких систем у різні галузі діяльності людини проблема забезпечення захисту і надійності їх функціонування стає все більш важливою і актуальною. До складу практично усіх сучасних мережевих комп'ютерних систем входить один або кілька серверів, які забезпечують інтеграцію з глобальною комп'ютерною мережею Internet. До типових функцій вебсерверів належить забезпечення сервісів WWW, FTP та електронної пошти.

Практичний досвід свідчить про неодноразові порушення працездатності мережевих комп'ютерних систем через порушення функціонування програмного забезпечення вебсервера внаслідок успішних кібератак типу «відмова в обслуговуванні» [2; 5; 6]. Очевидно, що попередження зазначених порушень пов'язано з розробкою ефективних засобів розпізнавання кібератак такого типу. На думку багатьох дослідників

[1; 12; 13; 17], підвищити ефективність таких коштів не можливо без розробки нових методів і моделей, на підставі яких можна сформулювати досить достовірний прогноз значень параметрів як в разі нормальних умов функціонування, так і при реалізації кібератак.

Тому безпека мережевих комп'ютерних систем безпосередньо залежить від достовірності методів і моделей прогнозу параметрів прогнозованості, розробка яких і є спільною метою цієї статті.

Аналіз останніх досліджень і публікацій. Базуючись на загальній методології розробки моделі оцінки параметрів захищеності веб-орієнтованих інформаційних систем, визначено [15], що архітектура такої моделі безпосередньо залежать від множини тих параметрів, які власне і визначають навантаження на вебсервер. Враховуючи типові задачі, які виконує інтернет-сервер, а також використовувані ним мережеві комунікаційні протоколи, визначено три групи параметрів, кожен з яких характеризує використання:

- апаратних ресурсів комп'ютера-сервера;
- ресурсів операційної системи;
- мережевих ресурсів.

Для визначення довгострокового прогнозу необхідно проаналізувати природу подій, їх можливий зв'язок з низкою інших легко прогнозованих подій. Розглянемо роботу вебсервера вищого навчального закладу. Тут повинен бути зв'язок із розкладом занять, з ритмами (циклами) життя людей. Повинна прослідковуватися добова періодичність, тижнева, семестрова, річна. Необхідно мати дані спостережень не менш ніж за рік. Виявлення періодів і їхня стійкість – це вже ключ до прогнозу. Оскільки спостереження відбуваються у часі, то і отриману функцію від часу ми будемо розглядати як сигнал.

Практичний досвід і результати [7] вказують на те, що ефективність більшості вітчизняних корпоративних інформаційних систем безпосередньо залежить від якості функціонування вебсерверу, за допомогою якого забезпечуються відповіді на запити віддалених інтернет-клієнтів. Очевидно, що обчислювальні ресурси вебсерверу повинні відповідати його завантаженості, інакше вебсервер або не зможе адекватно забезпечувати потреби користувачів інформаційної системи, або резервування надлишкових потужностей призведе до економічних втрат. Висновки і результати, отримані в [10], вказують на невідповідність існуючих інформаційних технологій прогнозування навантаження вебсерверу сучасним вимогам.

У [8] доведено, що недостатня точність прогнозування здебільшого пов'язана з недосконалістю існуючих методів і моделей прогнозу, які базуються на спрощеному математичному апараті і не дозволяють адаптуватися до типових процесів зміни параметрів навантаження. З точки зору виконання заданих функцій завантаженість вебсерверу залежить від його спроможності обслуговувати

HTTP-запити віддалених користувачів корпоративної комп'ютерної мережі. У першому наближенні можна вважати, що вебсервер витрачає однаковий обсяг обчислювальних ресурсів на виконання кожного запиту. Тому навантаження серверу здебільшого залежатиме від кількості отриманих HTTP-запитів і встановлених HTTP-з'єднань. Однак такі операції можуть бути деталізовані відповідно до протоколу TCP/IP, який є основою HTTP.

Навантаження вебсерверу можна оцінити за допомогою величин функціональних параметрів, які відповідають обсягам запитів і з'єднань по протоколах TCP/IP та HTTP. Зазначимо, що практичний досвід і результати [8] вказують на складний вигляд процесу зміни вказаних параметрів. У [5; 7] доведено, що для вебсерверу типовими є як стаціонарні, так і нестационарні режими експлуатації з багато-періодичним характером зміни величин функціональних параметрів. Нестационарність режимів експлуатації виявляється у нестационарності моментів виникнення / зникнення періодичних складників.

На рис. 1, 2 показано коливання графіків вхідного-вихідного трафіку та рівня навантаженості на прикладі досліджуваного вебсервера протягом тижня.

У результаті проведеного аналізу [15] можна стверджувати, що застосування вейвлет-перетворень є одним із найперспективніших напрямів підвищення ефективності систем прогнозування навантаження на вебсервер. Можна стверджувати, що в сучасних вейвлет-системах прогнозування не повністю відображені питання, пов'язані з вибором параметрів материнського вейвлету. При цьому теорія вейвлет-перетворень свідчить, що неправильний вибір цих параметрів може значно знизити ефективність системи прогнозування. Тому метою цієї роботи є створення методу вибору типу базисного вейвлету, найбільш ефективного при прогнозуванні параметрів навантаження на вебсервер.

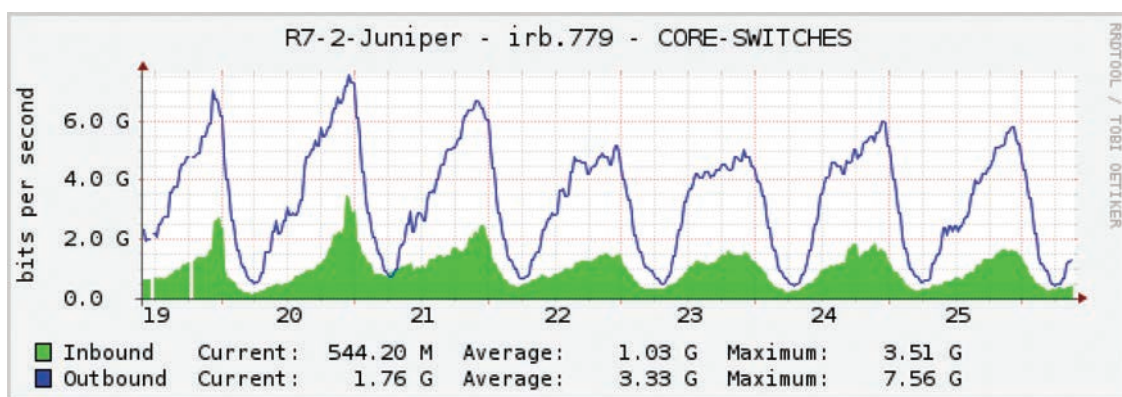


Рис. 1. Коливання вхідного та вихідного трафіку за тиждень на прикладі досліджуваного вебсервера

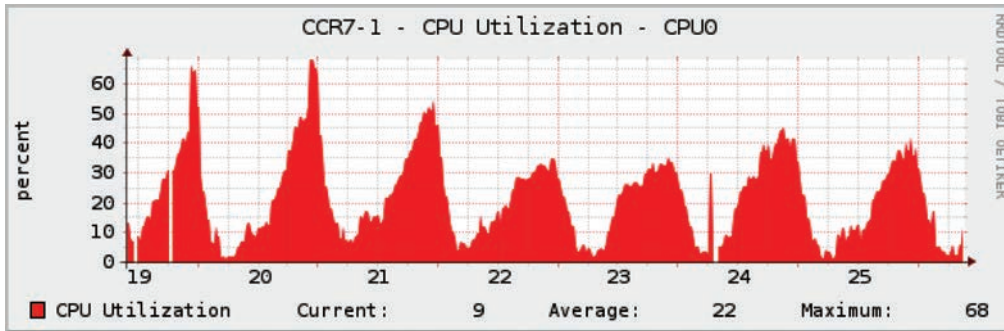


Рис. 2. Рівень навантаженості на сервер протягом тижня на прикладі досліджуваного вебсерверу

Постановка завдання. За аналогією з відомими методами розробки засобів прогнозування параметрів комп’ютерних систем [1; 2; 14; 16] розробляється метод вибору типу базисного вейвлету, який базується на таких положеннях:

1) критерії вибору найбільш ефективного типу базисного вейвлету повинні відображати міру його пристосованості до поставленої задачі – створення шаблону нормальної поведінки;

2) під k -им критерієм визначення найбільш ефективного типу базисного вейвлету будемо розуміти міру забезпечення в цьому типі вейвлету k -ої вимоги задачі створення шаблону;

3) розрахунок ефективності i -го типу базисного вейвлету можна представити у вигляді:

$$R_i = \sum_{k=1}^K \alpha_k r_k(i), \quad (1)$$

де R_i – показник інтегральної ефективності i -го типу базисного вейвлету, $r_k(i)$ – k -ий критерій ефективності i -го типу материнського вейвлету, α_k – ваговий коефіцієнт k -го критерію ефективності, K – кількість критеріїв ефективності;

4) найбільш ефективний тип базового вейвлету визначається так:

$$R_{eff} = \max \{R\}_I, \quad (2)$$

де $\{R\}_I$ – множина показників інтегральної ефективності типів базисних вейвлетів, I – потужність множини.

Формування базових положень розробки методу вибору типу базисного вейвлету дозволило перейти до наступного етапу досліджень – формування множини критеріїв ефективності, які використовуються у виразі 1). Для цього було проведено аналіз характеристик апробованих типів базисних вейвлетів (табл. 1).

Базою для аналізу слугували теоретичні роботи [7; 8; 10]. Аналіз проводився з точки зору вейвлет-перетворення дискретних сигналів, визначених на

Таблиця 1

Апробовані типи базисних вейвлетів

Назва	Аналітичний запис
WAVE-вейвлет	$\psi(t) = -t \exp(-\frac{t^2}{2})$
МНАТ	$\psi(t) = (1 - t^2) \exp(-\frac{t^2}{2})$
Гауссiан n -го порядку	$\psi_n(t) = (-1)^n \frac{\partial^n}{\partial t^n} (\exp(-\frac{t^2}{2}))$
DOG-вейвлет	$\psi(t) = \exp(-\frac{ t ^2}{2}) - \frac{1}{2} \exp(-\frac{ t ^2}{8})$
LP-вейвлет	$\psi(t) = (\pi t)^{-1} (\sin(2\pi t) - \sin(\pi t))$
Daubechies вейвлет	$\psi(t) = \sqrt{2} \sum_k g_k \varphi(2t - k)$
HAAR	$\psi(t) \geq \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & t < 0, t \geq 1 \end{cases}$
FНАТ	$\psi(t) \geq \begin{cases} 1, & t \leq 1/3 \\ -1/2, & 1/3 < t \leq 1 \\ 0, & t > 1 \end{cases}$
Morlet	$\psi(t) = \exp(ik_0 t) \exp(-\frac{t^2}{2})$
Paul	$\psi(t) = \Gamma(n+1) \frac{i^n}{(1-it)^{n+1}}$

кінцевому інтервалі часу. В результаті аналізу у першому наближенні було сформовано множину критеріїв ефективності:

rI – наявність у вейвлет-коефіцієнтах надлишкової інформації;

r_2 – відсутність у вейвлет-коефіцієнтах надлишкової інформації;

r_3 – можливість реалізації швидкого вейвлет-перетворення;

r_4 – наявність нескінченної регулярності;

r_5 – наявність випадкової регулярності;

r_6 – симетричність базисної функції;

r_7 – асиметричність базисної функції;

r_8 – ортогональність базисної функції;

r_9 – наявність масштабування функції;

r_{10} – можливість повного відновлення сигналу;

r_{11} – компактність базисної функції;

r_{12} – схожість геометрії базисної функції з геометрією аналізованого процесу.

За аналогією з [14; 16] прийнято, що в першому наближенні значення запропонованих критеріїв від r_1 до r_{11} можуть бути оцінені по двобальній дискретній шкалі. При цьому для i -го типу базисного вейвлету значення k -го критерію дорівнює

1, якщо відповідна k -та вимога повністю забезпечується, і дорівнює 0, якщо не забезпечується. Розраховані значення критеріїв для апробованих типів базисних вейвлетів наведені у табл. 2.

Виклад основного матеріалу дослідження. Використання розроблених базових положень визначення найбільш ефективного виду базисного вейвлету і результати робіт, присвячених побудові ефективних нейромережових моделей оцінки параметрів, дозволили запропонувати метод вибору типу базисного вейвлету, призначеного для розробки шаблону нормальної поведінки вебсерверу:

1) формалізація умов завдання розробки шаблону з метою визначення виду і значень параметрів, які визначають захищеність вебсерверу;

2) визначення значущості кожного із критеріїв ефективності, представлених у табл. 2. Визначення можливо реалізувати шляхом експертного оцінювання;

Таблиця 2

Значення критеріїв ефективності для апробованих типів базисних вейвлетів

Тип вейвлету	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}
WAVE	0	1	0	1	0	1	0	0	0	1	0
MHAT	0	1	0	1	0	1	0	0	0	1	0
Гаусіан	0	1	0	0	1	0	1	0	0	0	0
DOG	0	1	0	0	1	0	1	0	0	1	0
LP	0	1	0	0	1	0	1	0	0	0	0
Добеші	1	0	1	0	1	0	1	1	1	1	1
HAAR	1	0	1	0	1	0	1	1	1	1	1
FHAT	1	0	1	0	1	1	0	0	0	1	0
Morlet	0	1	0	1	0	1	0	0	0	1	0
Paul	0	1	0	1	0	1	0	0	0	1	0

Таблиця 3

Значення критерію r_{12} для апробованих типів базисних вейвлетів

Тип базису	WAVE	MHAT	Гаусіан	DOG	LP
r_{12}	0,4	0,5	0,5	0,5	0,5
Тип базису	FHAT	HAAR	Daubechies-5	Morlet	Paul
r_{12}	0,5	0,3	0,9	0,3	0,5

Таблиця 4

Вагові коефіцієнти критеріїв ефективності типу базисного вейвлету

r_1	r_2	r_3	r_4	r_5	r_6
0,15	0,01	0,07	0,06	0,06	0,06
r_7	r_8	r_9	r_{10}	r_{11}	r_{12}
0,06	0,06	0,06	0,15	0,06	0,2

Таблиця 5

Значення інтегрального показника ефективності R для апробованих типів базисних вейвлетів

Тип базису	WAVE	MHAT	Гаусіан	DOG	LP
R	0,36	0,38	0,23	0,38	0,23
Тип базису	FHAT	HAAR	Daubechies-5	Morlet	Paul
R	0,59	0,73	0,87	0,34	0,38

3) розрахунок ефективності кожного типу базисного вейвлету. Для цього необхідно використати вираз 1);

4) визначення за допомогою виразу 2) найбільш ефективного типу базисного вейвлету.

Розглянемо запропонований метод на конкретному прикладі вибору типу базисного вейвлету, призначеного для прогнозування навантаження на вебсервер:

1. Визначено, що шаблон нормальної поведінки вебсерверу необхідно оцінити на основі статистики звернень, графік якої наведено на рис. 1, 2. Таким чином прогнозована функція описується виразом $K=f(t)$, де K – кількість звернень до вебсерверу протягом доби, t – час.

2. Для кожного із апробованих типів базисних вейвлетів значення критерію $r12$ отримано в результаті порівняння геометрії відповідної функції з графіком статистики кількості звернень, показаним на рис. 1. Отримані значення наведені в табл. 3.

3. Оцінка важливості кожного з критеріїв ефективності у поставленій задачі була реалізована за допомогою експертного методу парного порівняння. Отримані результати показані в табл. 4.

4. Розрахунок показника інтегральної ефективності кожного із апробованих типів базисних вейвлетів реалізовано шляхом підстановки даних табл. 2-4 у вираз 1). Отримані результати наведені в табл. 5.

5. З використанням виразу 2) визначено, що найбільш ефективним є вейвлет Daubechies-5.

Визначення найбільш ефективного типу базисного вейвлету дозволило провести числові експерименти, спрямовані на верифікацію отриманих рішень шляхом перевірки точності апроксимації статистичних даних і можливості використовувати вейвлет-коефіцієнти для визначення локальних особливостей процесу навантаження. Розрахунки виконувалися у програмному середовищі Wavelet Toolbox системи MATLAB. У результаті аналізу були одержані графіки вейвлет-спектру, представлені на рис. 3.

Аналіз даних рис. 3 вказує на те, що в досліджуваному процесі простежуються 4 періодичні складники, які відповідають певному часу доби. Також проведено зворотне дискретне вейвлет-перетворення сигналу

по набору вейвлет-коефіцієнтів. Графіки відновленого і вихідного сигналу показані на рис. 4.

Можна вказати на досить високу схожість зазначених графіків. При цьому середня помилка апроксимації не перевищує 5%. У той же час середня помилка апроксимації за допомогою інших типів базисних вейвлетів склала близько 10%. Як показано у роботах [6; 12; 13; 17] помилка апроксимації у подібній статистиці знаходиться на рівні 5-7%.

Результати експериментів підтверджують можливість застосування розробленого методу, перспективи вдосконалення якого полягають в уточненні розрахунку критеріїв ефективності і їх вагових коефіцієнтів. Крім того, викликає інтерес деталізація запропонованого методу в напрямі розрахунку параметрів базисного вейвлету.

Висновки. Показано, що одним із найбільш перспективних напрямів розвитку систем прогнозування навантаженості на вебсервер є удосконалення математичного забезпечення шаблонів нормальної поведінки за рахунок застосування сучасних методів частотно-часового аналізу сигналів на базі теорії вейвлет-перетворень.

Визначено, що труднощі застосування теорії вейвлет-перетворень насамперед пов'язані з вибором типу базисного вейвлету, параметри якого мають бути адаптовані до умов застосування в конкретній системі розпізнавання кібератак.

Запропоновано метод вибору типу базисного вейвлету, найбільш ефективного при прогнозуванні параметрів навантаження на вебсервер. Метод базується на низці запропонованих положень і критеріїв ефективності, які дозволяють забезпечити при виборі типу базисного вейвлету

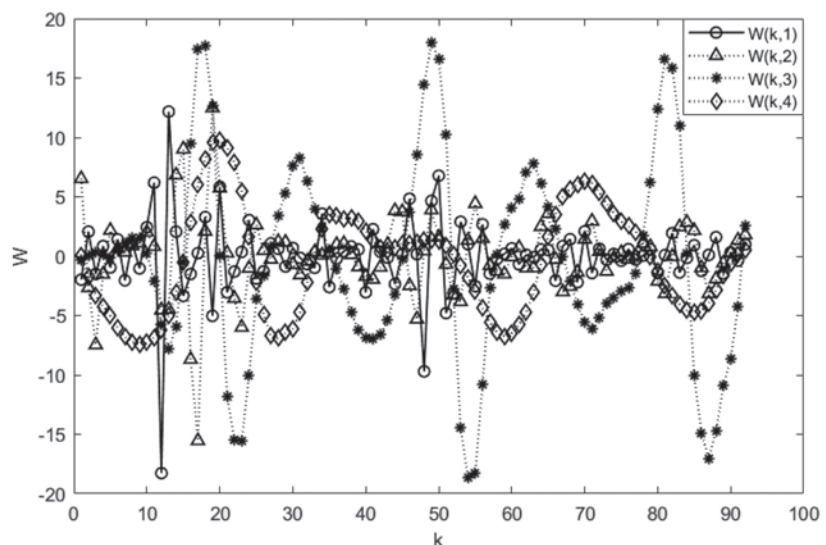


Рис. 3. Вейвлет-спектр відвідуваності вебсерверу

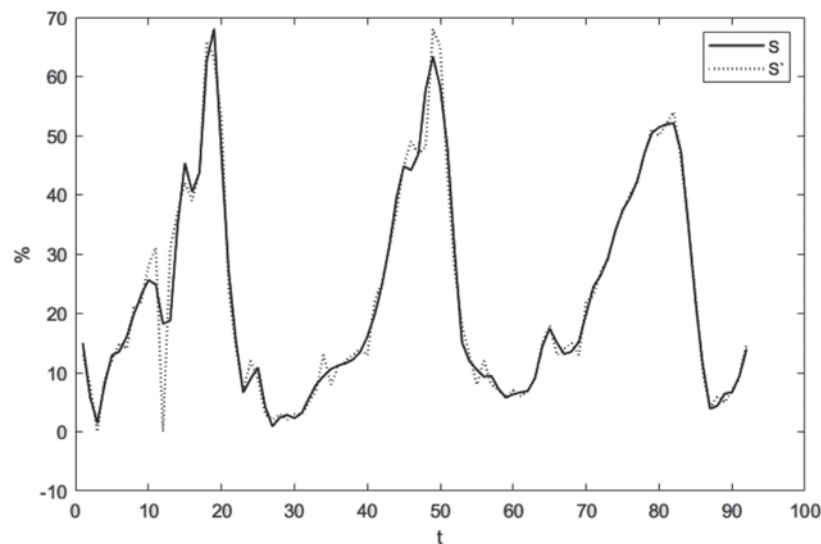


Рис. 4. Графіки вихідного (S) і відновленого сигналу (S')

важливі вимоги до поставленої задачі з метою розробки шаблону поведінки. При цьому шляхом численних експериментів з використанням зібраної

запропонованого методу за рахунок розробки процедури розрахунку параметрів базисного вейвлету.

них статистичних даних кількості звернень до вебсерверу одного з університетів України показано, що метод дозволяє обрати тип базисного вейвлету, використання якого забезпечує похибку апроксимації на рівні похибки сучасних систем прогнозування навантаження вебсерверів. Важливою перевагою методу є можливість уникнення в процесі розробки вейвлет-моделі тривалих і численних експериментів, спрямованих на визначення типу базисного вейвлету.

Перспективи подальших досліджень пов'язані з уточненням процесу розрахунку критеріїв ефективності та удосконаленням

Список літератури:

1. Bariyev I.M., Aitchanov B.H., Tereikovskiy I.A., Tereikovska L.A., Korchenko A.A. (2017). "Deep neural networks in cyber attack detection systems", International Journal of Civil Engineering and Technology (IJCIET), Volume 8, Issue 11, November 2017, p. 1086–1092.
2. Довлад О.А. Дослідження та розробка моделі процесу атаки та трафіку локальної мережі / О.А. Довлад // Захист інформації. 2009. № 1. Р. 83–86.
3. Dychka I., Tereikovskiy I., Tereikovska L., Pogorelov V., Mussiraliyeva S. (2018). "Deobfuscation of computer virus malware code with value state dependence graph". Advances in Intelligent Systems and Computing, p. 370–379, 2018. DOI: 10.1007/978-3-319-91008-6.
4. Менаске Д. Производительность Web-служб. Анализ, оценка и планирование / Менаске Д., Виргилио А.; пер. с англ. СПб: «ДиаСофтЮп», 2003, 480 с.
5. Переберин А.В. О классификации вейвлет-преобразований. Выч. мет. программирование, 2:3 (2001), с. 15–40.
6. Qi-Song Chen; Xin Zhang; Shi-Huan Xiong; Xiao-Wei Chen Short-term power load forecasting with least squares support vector machines and wavelet transform. 2008 International Conference on Machine Learning and Cybernetics Year: 2008, Volume: 3. Pages: 1425–1429.
7. Sitnikov V.S., Bilenko A.A. Classification of wavelet functions. Труды Одесского политехнического университета, 2008, вып. 1(29). С. 168–171.
8. Сыропятов А.А. Метод мониторинга трафика защищенных высокоскоростных коммерческих сетей нового поколения / А.А. Сыропятов // Наукові записки УНДІЗ. 2009. № 2(1). С. 65–73.
9. Steinbuch M. Wavelet Theory and Applications A literature study / M. Steinbuch, M.J.G. van de Molengraft. Eindhoven University of Technology, 2005. 39 с.
10. Shuping Yao; Changzhen Hu; Wu Peng. Server Load Prediction Based on Wavelet Packet and Support Vector Regression 2006. International Conference on Computational Intelligence and Security Year: 2006, Volume: 2 Pages: 1016–1019.
11. Tereykovska L., Tereykovskiy I., Aytkhozhayeva E., Tynymbayev S., Imanbayev A. (2017). "Encoding of neural network model exit signal, that is devoted for distinction of graphical images in biometric authenticate systems", News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences, Volume 6, Number 426, p. 217–224, 2017.
12. Zhengbing Hu, Igor A. Tereykovskiy, Lyudmila O. Tereykovska, Volodymyr V. Pogorelov. "Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems", International Journal of Intelligent Systems and Applications (IJISA), Vol. 9, № 10, p. 57–62, 2017. DOI: 10.5815/ijisa.2017.10.07.

13. Zhengbing Hu, Igor A. Tereykovskiy, Lyudmila O. Tereykovska, Mykola M. Tsiutsiura, Kostiantyn O. Radchenko. Applying Wavelet Transforms for Web Server Load Forecasting // Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing, vol. 938, p. 13–22 (2019). Springer. Cham: https://doi.org/10.1007/978-3-030-16621-2_2.

14. Zijiang Yang Research on Server Load Prediction Based on Wavelet Packet Theory 2007. First IEEE International Symposium on Information Technologies and Applications in Education Year: 2007 Pages: 610–613.

15. Якубен М.Б. Обнаружение сетевых атак методом поиска аномалий на основе вероятностного и верификационного моделирования / М.Б. Якубен // Штучний інтелект. 2005. № 3. С. 679–687.

Radchenko K.O. CONCEPTUAL MODEL FOR ENSURING EFFICIENCY OF WEB SERVER LOAD FORECASTING

The subject of the article is related to the solution of the current scientific and practical problem of improving the mathematical software for forecasting the load on the web server. As a result of the analysis of literary sources, the prospect of improving the patterns of normal behavior of web servers was determined by introducing in their mathematical support of modern methods of time-frequency signal analysis based on the theory of wavelet transforms.

The lack of illumination in the available literature of issues related to the justification of adapting the type of the basic wavelet with the conditions for creating a pattern of the normal behavior of a web server was also determined. It is proposed to correct this drawback by creating a method for determining the effective type of the basic wavelet, which is designed to develop a pattern of normal behavior. For this, a number of provisions and a list of performance criteria are justified, which allow for the selection of the effective type of the basic wavelet in accordance with the significant requirements of the development task of the specified template.

The essence of the method consists in determining the parameters characterizing the basic wavelets and their correlation with the important conditions for the formation of the pattern of normal behavior of the web server. Using computer experiments, it was shown that the method allows choosing the type of basic wavelet that takes into account in the constructed pattern the normal behavior of the complex multi-period nature of the security parameters of the web server. The accuracy of this pattern is at the level of modern web server load forecasting systems. At the same time, an important advantage of the method is the avoidance during the development of the wavelet model of the normal behavior pattern of long-term numerous experiments aimed at determining the type of the base wavelet.

Prospects for further research are associated with the development of the proposed method in the direction of adapting the parameters of the basic wavelet to the conditions of the task.

Key words: *Wavelet model, pattern of normal behavior, web server, information protection, base wavelet.*