

УДК 004.89+519.216.3+656.2

Т. О. САВЧУК, А. В. КОЗАЧУК

Вінницький національний технічний університет

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ СТАНУ ТЕХНОГЕННОЇ СИТУАЦІЇ

Анотація. В статті описані складові інформаційної технології прогнозування стану техногенної ситуації, проведено аналіз результатів використання цієї технології для прогнозування кількості втрачених пакетів у локальній мережі, а також проведено аналіз адекватності математичної моделі, що входить до складу запропонованої інформаційної технології.

Ключові слова: інформаційна технологія, імітаційне моделювання, багатовимірне представлення даних.

Аннотация. В статье описаны составляющие информационной технологии прогнозирования состояния техногенной ситуации, проведен анализ результатов использования этой технологии для прогнозирования количества потерянных пакетов в локальной сети, а также проведен анализ адекватности математической модели, входящей в состав предложенной информационной технологии.

Ключевые слова: информационная технология, имитационное моделирование, многомерное представление данных.

Abstract. The paper describes the components of information technology of forecasting of technological situation and analysis of results of usage of this technology to predict the number of lost packets on the local network, as well as an analysis of the adequacy of mathematical model, which is a part of the proposed information technology.

Key words: information technology, simulation, multi-dimensional data representation.

Вступ

Прогнозування стану техногенної ситуації є комплексним процесом, що складається з декількох етапів на яких відбувається уточнення типу техногенної ситуації, що визначає метод прогнозування стану техногенної ситуації та формування подальших рекомендацій. Перераховані етапи з'єднуються між собою інформаційними потоками, що містять вхідні та вихідні дані кожного з етапів прогнозування стану техногенної ситуації. Для систематизації процесу прогнозування стану техногенної ситуації етапи цього процесу доцільно об'єднати у технологічний ланцюжок, що дозволяє розглядати їх як інформаційну технологію [1]

Розробка інформаційної технології прогнозування стану техногенної ситуації

У відповідності з класичним визначенням [2, 3] інформаційної технології, інформаційну технологію прогнозування стану техногенної ситуації можна представити як сукупність процесів визначення значимих характеристик техногенної ситуації та пошуку кореляції між ними, внесення початкових даних про стан техногенної ситуації, вибору метода прогнозування стану техногенної ситуації та безпосереднього прогнозування її стану за допомогою моделювання гіперкубу станів техногенної ситуації або інших методів інтелектуального аналізу даних, формування рекомендацій щодо дій стосовно техногенної ситуації і програмного забезпечення прогнозування стану техногенної ситуації, об'єднані в технологічний ланцюжок з метою підвищення ефективності прогнозування стану техногенної ситуації.

Інформаційна технологія прогнозування стану техногенної ситуації складається з наступних етапів:

1. Попередня обробка інформації про техногенну ситуацію.

- a. Визначення набору значимих характеристик техногенної ситуації – виділення властивостей техногенної ситуації, можливі значення яких є відомим та таких, що можуть бути використаними для прогнозування стану техногенної ситуації.
- b. Внесення початкових даних про стан техногенної ситуації – впорядкування наявної інформації про можливі стани техногенної ситуації та внесення її у сховище даних.
- c. Пошук кореляції між характеристиками техногенної ситуації – необхідно для збільшення швидкодії алгоритмів прогнозування за рахунок зменшення розмірів структур даних.
- d. Вибір метода прогнозування стану техногенної ситуації – застосування критеріїв доцільності використання багатовимірного підходу до прогнозування стану техногенної ситуації. На основі вибору виконується один або декілька перелічених кроків:
- e. Побудова та прогнозування стану гіперкубу
- f. Підготовка даних для проведення навчання алгоритмів «Дерева прийняття рішень», «Нейронна мережа», «Наївний Баєс»

2. Прогнозування стану техногенної ситуації

1. Прогнозування за допомогою гіперкубу станів техногенної ситуації
2. Прогнозування за допомогою дерев прийняття рішень та нейронних мереж
3. Формування рекомендацій щодо дій стосовно техногенної ситуації.

Основні етапи інформаційної технології прогнозування стану техногенної ситуації зображена на рис. 1.

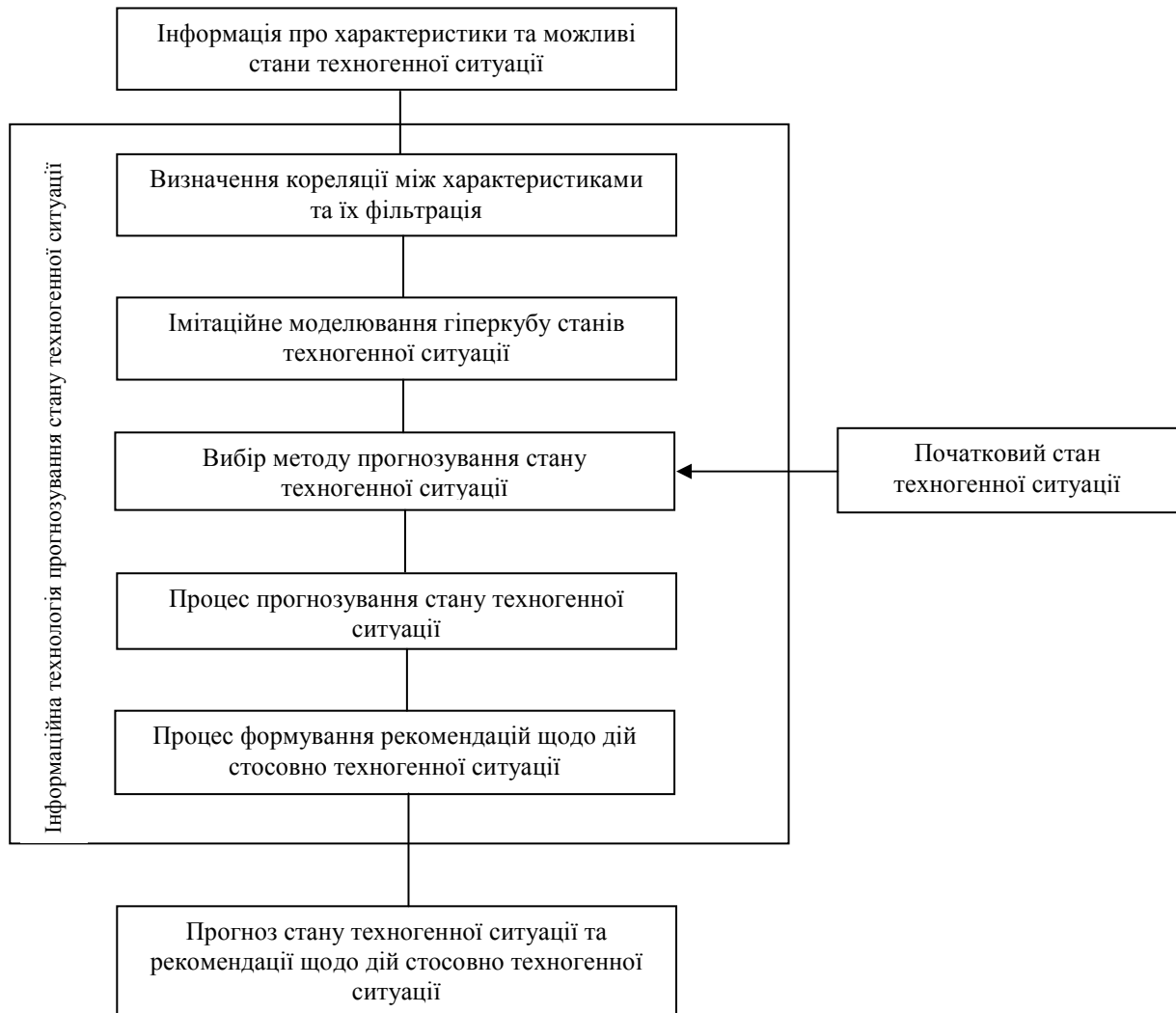


Рисунок 1 – Основні етапи інформаційної технології прогнозування стану техногенної ситуації

Аналіз результатів застосування

Дослідимо результати застосування інформаційної технології прогнозування стану техногенної ситуації для прогнозування відсотку втрачених кадрів маршрутизатором D-Link Dir300, що працює в режимі комутатора у ТОВ «СДМ Україна». Кількість втрачених кадрів залежить від таких характеристик використання комутатора як середній розмір пакета, розподілення трафіку між портами комутатора, обсяг пам'яті черги комутатора, внутрішня архітектура комутатора та максимальний розмір адресної таблиці. При цьому під час роботи комутатора можуть змінюватися такі характеристики як середній розмір кадру та трафік між портами комутатора. Модель D-Link Dir300 має чотири порти. З кожного порту може відбуватися передача інформації на будь-який з трьох інших. Таким чином, можна виділити $4 \cdot 3 = 12$ характеристик, що описують техногенну ситуації. Ще однією характеристикою є середній розмір кадру [4, 5]. Обчислення проводиться на комп'ютері з двох ядерним процесором з частотою кожного ядра 2.4 ГГц та обсягом оперативної пам'яті 4 ГБ.

Таким чином, позначивши за tr_{ij} трафік між портами i та j , можна визначити вектор характеристик техногенної ситуації W :

$$W = \{\text{середній розмір кадру; } t_{12}; t_{13}; t_{14}; t_{21}; t_{23}; t_{24}; t_{31}; t_{32}; t_{34}; t_{41}; t_{42}; t_{43}; \text{ відсоток втрачених кадрів}\}.$$

Потужність вектору W рівна 14:

$$n = 14.$$

Цільовою характеристикою техногенної ситуації, значення якої прогнозується є w_{14} – відсоток втрачених кадрів:

$$w_m = w_{14}.$$

Вектор V характеристик техногенної ситуації, що використовуються при прогнозуванні значення цільової характеристики має потужність $n-1 = 13$ та складатиметься з таких характеристик використання комутатора:

$$V = \{\text{середній розмір кадра; } t_{12}; t_{13}; t_{14}; t_{21}; t_{23}; t_{24}; t_{31}; t_{32}; t_{34}; t_{41}; t_{42}; t_{43}\}$$

Для досягнення прийнятної складності обчислень при прогнозуванні відсотку втрачених комутатором кадрів за допомогою інформаційної технології прогнозування стану техногенної ситуації, можливі діапазони значень характеристик техногенної ситуації повинні бути розбиті на групи, що описують можливі значення характеристики.

Розмір кадра було розбито на дві групи: малий (64 – 256 байт) та великий (257 – 1518 байт).

Інформацію про трафік між портами у чотирьохпортового комутатора, що використовує протокол Fast Ethernet розбито на групи в залежності від кількості інформації, що передається в одному напрямі між парою портів: слабкий (0 – 33 Мб/с), середній (34 – 66 Мб/с) та сильний (67 – 100 Мб/с).

Таким чином, гіперкуб станів техногенної ситуації, що використовується для прогнозування відсотку втрачених комутатором кадрів, містить $3^{12} * 2 = 1062882$ комірок. Така кількість комірок дозволяє проводити моделювання стану гіперкуба на персональному комп’ютері.

Для використання методу прогнозування стану техногенної ситуації за допомогою гіперкуба станів техногенної ситуації необхідно також визначити коефіцієнти c_i , що описують вплив характеристики w_i на цільову характеристику v_m . Комутатор працює у напівдуплексному режимі, при якому кожен порт може подавати інформацію або на вхід або на вихід в конкретний момент часу, що означає однаковий вплив як вхідного так і вихідного трафіка порта на кількість втрачених кадрів. Враховуючи рівнозначність всіх портів комутатора, вплив кожного з портів на кількість втрачених кадрів є рівним і не залежним від напрямку передачі даних. Розмір кадру впливає на кількість втрачених кадрів, визначаючи співвідношення корисної і службової інформації у трафіку. Загальний вплив розміру кадрів суттєво менший за вплив завантаженості каналів. Враховуючи вищезазначені твердження, можна ввести наступні значення коефіцієнтів c_i :

$$c_1 = 10; c_2 = c_3 = c_4 = \dots = c_{13} = 50.$$

Враховуючи симетричність внутрішньої будови комутатора та той факт, що кількість втрачених кадрів залежить в від сумарного потоку інформації, що надходить на кожен з портів комутатора, можна ввести альтернативний набір характеристик V_1 та розрахувати їх значення:

$$V_1 = \{\text{середній розмір кадра; } t_{вх1}; t_{вх2}; t_{вх3}; t_{вх4}; t_{вих1}; t_{вих2}; t_{вих3}; t_{вих4}\},$$

де $t_{вхi}$ – сумарний трафік, що надходить на i -й порт з інших портів комутатора, $t_{вихi}$ – сумарний трафік, що надходить на i -й порт комутатора ззовні. Значення характеристик вектору V_1 можуть бути розраховані наступним чином:

$$v(t_{вхi}) = v(t_{1i}) + v(t_{2i}) + v(t_{3i}) + v(t_{4i}), \quad (4.1)$$

$$v(t_{вихi}) = v(t_{i1}) + v(t_{i2}) + v(t_{i3}) + v(t_{i4}).$$

Вектор характеристик V_1 має меншу потужність у порівнянні з вектором V і при цьому містить таку ж кількість корисної інформації про техногенну ситуацію, що досягнуто за допомогою використання додаткових знань про предметну область, а саме – знання про факт того, що кількість пакетів, що втрачаються на порту залежить від сумарного трафіку через цей порт.

Враховуючи факт, що через особливості роботи комутатора $t_{ii} = 0$ та виходячи із співвідношення 4.1, значення величин $t_{вхi}$, можуть змінюватись в діапазоні 0 – 300 Мб/с [6], варто змінити діапазони групування цих величин на наступні: слабкий (0 – 100 Мб/с), середній (101 – 200 Мб/с) та сильний (201 – 300 Мб/с).

Варто зазначити, що внутрішній трафік маршрутизатора може перевищувати зовнішній у $m-1$ разів, де m – кількість портів комутатора. Таке збільшення трафіка зумовлене відсутністю у таблицях комутації MAC-адрес, які зазначені як адреси отримувачів пакетів, що проходять через комутатор [53]. При звичайних умовах експлуатації внутрішній трафік комутатора незначно перевищує зовнішній, це означає, що значна кількість комірок гіперкубу станів техногенної ситуації містить комірки, що з малою імовірністю будуть зустрічатися при нормальних умовах експлуатації гіперкуба. Кількість таких комірок складає 9664. Також існують повністю невизначені комірки, значення яких не може бути отримане експериментально, кількість таких комірок – 846. Загальна кількість комірок гіперкубу станів техногенної ситуації – 13122.

Для застосування вектору V_1 необхідно також ввести нові значення коефіцієнтів відстані c_{1i} . Так як природа характеристик вектору V_1 така ж, як і у вектору V , значення коефіцієнтів c_{1i} змінювати не потрібно, зміниться тільки їх кількість:

$$c_{1i} = c_i, i \in [1;9].$$

Розрахуємо значення критеріїв доцільності використання багатовимірного підходу. Експериментальним шляхом було отримано початкове значення 7442 комірок гіперкубу, з них 2610 комірок описують стани комутатора, що можуть виникнути при повсякденному його використанні.

Значення базового критерія доцільності використання багатовимірного підходу до прогнозування техногенної ситуації рівне

$$Q_6 = \frac{N_{u6}}{N} = \frac{7442}{13122} = 0.567.$$

Для розрахунку значення зваженого критерія доцільності використання багатовимірного підходу до прогнозування техногенної ситуації потрібно визначити величину N_{uz} ,

$$N_{uz} = \sum_{k_1=1}^{L_1} \sum_{k_2=1}^{L_2} \dots \sum_{k_{n-1}=1}^{L_{n-1}} W_{k_1, k_2, \dots, k_{n-1}} \alpha_{k_1, k_2, \dots, k_{n-1}}.$$

Прийнявши значення коефіцієнта w за 1 для комірок, що описують стан комутатора, що може бути досягнутий за нормальних умов експлуатації, та за 0.05 для комірок, що описують стан комутатора, що малоймовірно може бути досягнутий за нормальних умов експлуатації, отримуємо:

$$N_{uz} = 1 * 260 + (7442 - 2610) * 0.05 = 2851,6.$$

Відповідне значення критерія доцільності використання багатовимірного підходу до прогнозування техногенної ситуації рівне

$$Q_3 = \frac{N_{uz}}{N} = \frac{2851,6}{13122} = 0,217.$$

Значення крок часу прогнозування Δt було обрано як 1 с. Таке значення кроку забезпечує достатню швидкодю та прийнятну точність прогнозування. Кількість кроків прогнозування P , таким чином, складає $250/1 = 250$ с.

Розрахунок значення зваженого критерія доцільності використання багатовимірного підходу до прогнозування техногенної ситуації з урахуванням кількості кроків моделювання Q_k є недоцільним, так як кількість кроків моделювання – 250 перевищує діаметр гіперкубу. За таких умов

$$Q_k = Q_3 = 0,217.$$

Конкретні умови експлуатації, для яких проводиться прогнозування заздалегідь невідомі, тому розрахувати значення гармонічного критерія доцільності використання багатовимірного підходу до прогнозування техногенної ситуації Q_7 неможливо.

Отже, на основі алгоритму вибору методу прогнозування стану техногенної ситуації слід використовувати зважений критерій доцільності використання багатовимірного підходу до прогнозування техногенної ситуації. Експертним шляхом було визначено мінімальне допустиме значення зваженого критерія доцільності використання багатовимірного підходу до прогнозування техногенної ситуації на рівні 0.15. Таким чином, для прогнозування кількості втрачених комутатором пакетів слід використовувати імітаційне моделювання гіперкубу станів техногенної ситуації.

Для визначення кількості втрачених комутатором пакетів також було застосовано метод прогнозування стану техногенної ситуації за допомогою регресійних дерев прийняття рішень. Для реалізації прогнозу було використана платформа Microsoft SQL Server Analysis Services. Експериментальні дані були записані в таблицю реляційної бази даних, кожен стовпець якої відображав одну характеристику техногенної ситуації. За допомогою DMX-запитів були створені багатовимірна модель зберігання даних (англ. Mining Structure) та модель аналізу даних (англ. Mining model), що базується на реалізації алгоритму дерев прийняття рішень під назвою «Microsoft Decision Trees Algorithm». Так як кількість характеристик техногенної ситуації – 9 не є значною, додаткових заходів щодо регулювання глибини дерева прийняття рішень не проводилось. Значення параметрів алгоритму за замовченням не змінювались.

Для наведених вище умов експлуатації було експериментально визначено кількість втрачених пакетів. Порівняння результатів прогнозування з двома методами наведено в табл. 1.

Таблиця 1 – Порівняння результатів прогнозування з результатом інформаційної технології прогнозування стану техногенної ситуації

Час, с	Кількість втрачених пакетів за 1/10 с		
	Експериментальні дані	Імітаційне моделювання	Дерева прийняття рішень
1	2	3	4
10	548	537,1624761	532,69
20	540	547,7793894	540,38
30	526	554,6942891	548,07
40	570	559,1159696	555,76

Продовження таблиці 1

1	2	3	4
50	551	561,8707638	563,45
60	560	563,5215559	571,14
70	569	564,4501175	573,82
80	560	564,913996	573,08
90	554	565,085767	572,34
100	561	565,0800752	571,6
110	560	564,9722274	570,86
120	564	564,8109515	570,12
130	567	564,6271275	569,38
140	560	564,4397463	568,64
150	553	564,2599593	567,9
160	551	564,0938191	567,16
170	539	563,9441233	566,42
180	567	563,8116459	565,68
190	562	563,6959522	564,94
200	563	563,5959314	564,2
210	570	563,5101382	563,46
220	565	563,4370081	562,72
230	565	563,3749874	561,98
240	534	563,3226078	561,24
250	550	563,2785255	560,5

Графічна інтерпретація отриманих результатів порівняння представлена на рис. 2.

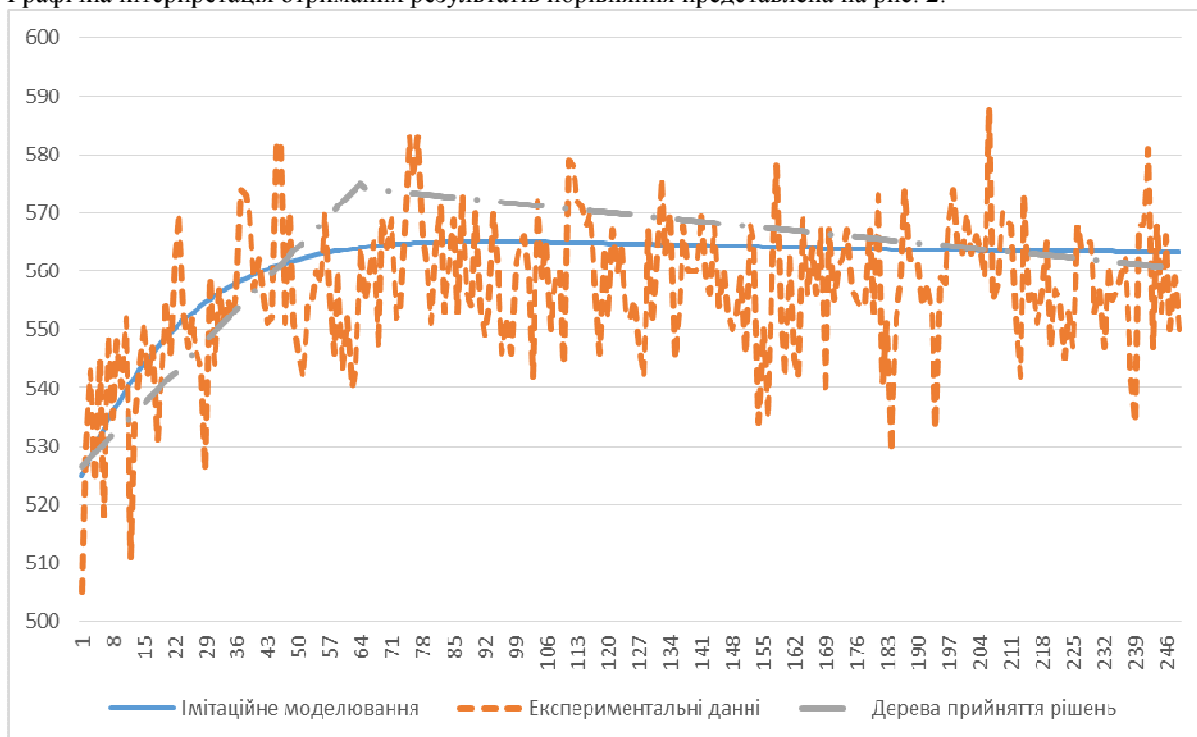


Рисунок 2 – Порівняння результатів прогнозування кількості втрачених пакетів, отриманих різними методами

На основі даних з таблиці 1 можна розрахувати середнє квадратичне відхилення результатів, отриманих методами імітаційного моделювання гіперкубу станів техногенної ситуації та регресійних дерев прийняття рішень. Значення середнього квадратичного відхилення для різних горизонтів прогнозування наведені у таблиці 2.

Таблиця 2 – Середнє квадратичне відхилення при прогнозуванні кількості втрачених пакетів

Кількість кроків прогнозування	СКВ результатів імітаційного моделювання	СКВ результатів дерев прийняття рішень
50	11,77	12,46
100	11,66	14,72
250	11,65	13,7

Таким чином, точність результату, отриманого за допомогою імітаційного моделювання гіперкубу старів техногенної ситуації вища від 7% (для коротких проміжків часу) до 21% (для середніх та довгих проміжків часу).

Оцінка адекватності моделі

Оцінимо адекватність моделі гіперкубу станів техногенної ситуації як ступінь відповідності моделі реальному об'єкту або процесу, для опису якого вона будується [8].

Модель є насиченою, так як потужність множини експериментальних даних $N=250$ перевищує кількість параметрів моделі $m=9$, що робить визначення адекватності моделі гіперкубу станів техногенної ситуації можливим.

Перевірка адекватності математичної моделі даними експерименту проводиться на основі зіставлення дисперсії відтворюваності середнього значення функції відгуку $s_2(y)$ і дисперсії адекватності. Оцінка дисперсії адекватності при $N > m$ характеризує відхилення між результатами спостережень і значеннями, формованими по функції відгуку.

Визначимо дисперсію адекватності використовуючи співвідношення

$$\sigma_a^2 = \frac{1}{N-m} \sum_{u=1}^N (y_u - y_u')^2,$$

де y_u – експериментальне значення кількості втрачених пакетів,

y_u' – прогноз кількості втрачених пакетів отриманих за допомогою гіперкуба станів техногенної ситуації [9]. Підставивши дані з таблиці 4.1 в наведене співвідношення отримаємо

$$\sigma_a^2 = \frac{1}{250-9} 33671,27 = 139,71$$

Визначимо дисперсію експериментальних даних:

$$\sigma_y^2 = \frac{1}{N} \sum_{u=1}^N (y_u - \bar{y})^2,$$

де \bar{y} – математичне сподівання експериментальних даних.

$$\bar{y} = \frac{1}{N} \sum_{u=1}^N y_u = 556,58$$

$$\sigma_y^2 = \frac{1}{250} 37164,56 = 148,65$$

На основі величин σ_y^2 та σ_a^2 розрахуємо експериментальне значення F-критерія Фішера [10]:

$$F = \text{MAX} \left(\frac{\sigma_a^2}{\sigma_y^2}; \frac{\sigma_y^2}{\sigma_a^2} \right) = \text{MAX}(0,9398; 1,0639) = 1,0639.$$

Модель гіперкубу станів техногенної ситуації можна вважати адекватною, якщо значення F менше за критичне, що визначається за таблицями F-розподілу. Для рівня значимості $p=0,05$ критичне значення критерія Фішера $F_{кр}$ складає 3,18. Так як $F < F_{кр}$, модель гіперкубу станів техногенної ситуації можна вважати адекватною.

Висновки

1. Розроблено інформаційну технологію прогнозування стану техногенної ситуації, що дозволяє на основі інформації про можливі стани та початкові параметри техногенної ситуації отримати прогноз її розвитку та рекомендації про подальші дії щодо техногенної ситуації.

2. Проведено оцінку адекватності математичної моделі процесу розвитку техногенної ситуації на основі застосування F-критерія Фішера. Для рівня значимості $p=0,05$ розраховане значення критерія Фішера менше за критичне, з чого можна зробити висновок про те, що модель – адекватна.

3. Досліджено результати впровадження інформаційної технології прогнозування стану техногенної ситуації на ТОВ «СДМ Україна». Показано, що кількість втрачених пакетів на 1/10 секунди може бути зпрогнозована з середнім квадратичним відхиленням менше 15.

Список літератури

1. Гвоздева В.А. Информатика, автоматизированные информационные технологии и системы // М.: Форум, – 2011г. – 544с. – ISBN 978-5-8199-0449-7.
2. Лихачева Г.Н. Информационные технологии: Учебно-практическое пособие / Г.Н. Лихачева, М.С. Гаспариан. // М.: Изд. центр ЕАОИ. – 2007. – 189 с. – ISBN 978-5-374-00032-0.
3. Корнеев И.К. Информационные технологии. Учебник / И.К. Корнеев, Г. И. Ксандопуло, В.А. Адамович. // Проспект, 2009. – 224с. - ISBN: 978-5-482-01401-1.
4. Ашигалиев Д.У. Анализ требований к производительности коммутаторов / Ашигалиев Д.У., Амиргалиев Е.Н., Сейлова Н.А.// г. Алматы, Республика Казахстан. Научный журнал МО и Н Поиск 2010-10-16 2010, №1
5. Устранение проблем интерфейса и неполадок портов коммутатора // Cisco Systems, Inc. Режим доступа: http://www.cisco.com/cisco/web/support/RU/10/105/105430_53.pdf
6. Kuan-Ta C. Toward an Understanding of the Processing Delay of Peer-to-Peer Relay Nodes/ Kuan-Ta Chen, Jing-Kai Lou// Proceedings of IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2008. Режим доступа: http://mmnet.iis.sinica.edu.tw/pub/chen08_relay.pdf
7. T. Ferrari QoS and Multiprotocol Label Switching Experiments for the Design of an ATM-Based National Network/ Tiziana Ferrari, Antonia Ghiselli, Cristina Vistoli// INET/IGC 2004 "Strengthening the Net: Building an Open and Trusted Internet"
8. Шаригін О. А. Розробка підходу до перевірки адекватності моделі прийняття рішень з нечіткими параметрами / О. А. Шаригін // Оптико-електронні інформаційно-енергетичні технології. – 2012. – № 1(23). – С. 59– 61. – ISSN 1681-7893.
9. Основы теории надежности и диагностика. Новгородский государственный университет имени Ярослава Мудрого. Режим доступа: <http://www.novsu.ru/npe/files/um/1128/umk/OTND/index.htm>
10. Красовский Г.И. Планирование эксперимента/ Красовский Г.И., Филаретов Г.Ф. // Минск: Изд-во БГУ им. В.И. Ленина, 1982. – 302 с.
Стаття надійшла: 19.06.2014.

Відомості про авторів

Савчук Тамара Олександрівна – к.т.н., проф., кафедри Комп'ютерних наук ВНТУ, м. Вінниця, Хмельницьке шосе 95.

Козачук Андрій Валерійович – аспірант кафедри Комп'ютерних наук ВНТУ. м. Вінниця, Хмельницьке шосе 95.