

5.2 Ідентифікаційний експеримент

Для складних реальних об'єктів в більшості випадків повністю чи частково відсутня *априорна інформація* про структуру об'єкта, і її потрібно отримати за тими самими реалізаціями вхідних і вихідних змінних, за якими визначаються параметри моделі об'єкта.

Джерелом інформації про об'єкт в будь-якому випадку є дослід (експеримент). Розрізняють два види дослідів: пасивний і активний.

Пасивним називається той експеримент, який проводиться шляхом спостереження змін стану об'єкта під дією зовнішніх впливів в процесі його нормального функціонування.

Активний експеримент передбачає генерування діючих на об'єкт *тестових сигналів* потрібної форми, що скорочує тривалість експерименту і спрощує наступне визначення параметрів моделі. Ці сигнали поділяються на регулярні і випадкові.

Часто для отримання даних про процес використовують *активно-пасивний реєстраційний експеримент*. Він полягає в тому, що коливання вносять лише по деяких каналах; при цьому не ставиться задача строгої стабілізації інших параметрів. Наявність елементів активного експерименту дозволяє розширити діапазон зміни параметрів процесу, що забезпечує більшу достовірність отриманих даних.

Деякі моделі можуть бути отримані еволюційними методами. Такі методи використовуються при моделюванні складних систем, для яких відомі закони розвитку і початковий стан. Еволюційне моделювання близьке за змістом до прогнозування, але відрізняється дуже великою розмірністю задачі. Такий спосіб моделювання незамінний при синтезі і ідентифікації моделей замкнених екосистем, космічних об'єктів, розповсюдження епідемій тощо. Однією з найбільших проблем еволюційного моделювання є великий обсяг розрахунків. Для його реалізації використовуються найпотужніші комп'ютери, а у відповідальних випадках – суперкомп'ютери.

5.2.1 Пасивна ідентифікація функціональної моделі

Схема пасивного експерименту наведена на рис. 5.2, а.

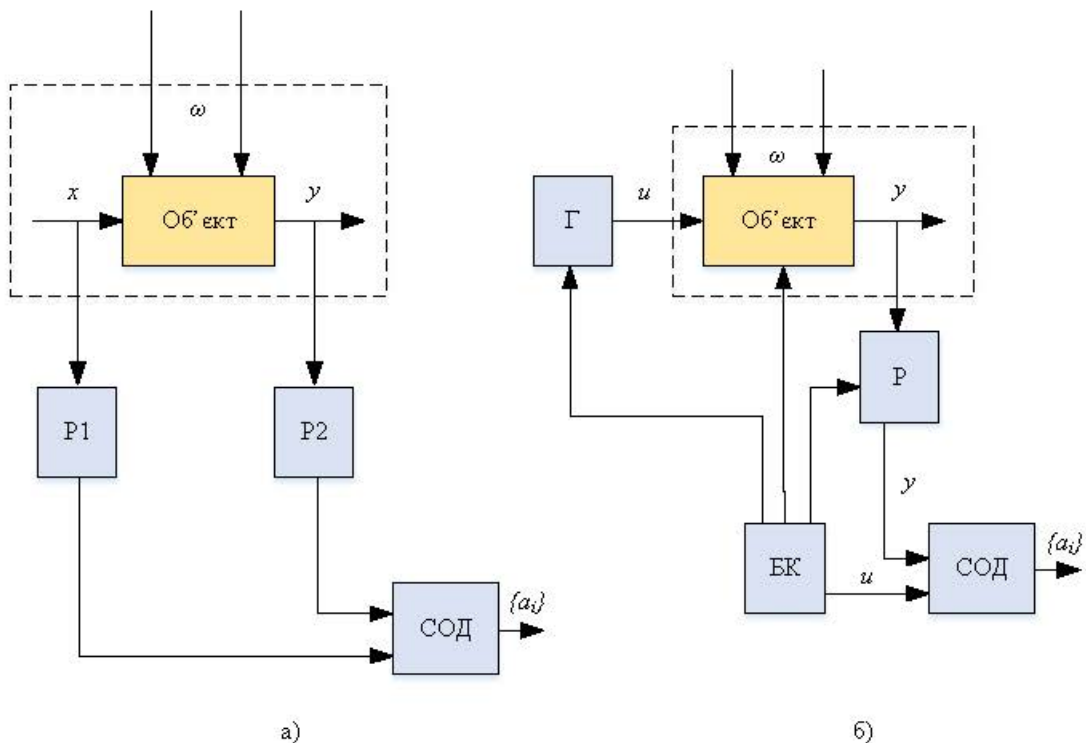


Рисунок 5.2 – Схеми експериментів: а) пасивний експеримент,
б) активний експеримент

Р – реєстратор, СОД – система обробки даних, Г – генератор тестових сигналів, БК – блок керування, $\{a_i\}$ – вектор шуканих параметрів

Перевагою пасивного методу ідентифікації є відсутність порушень природного режиму функціонування об'єкта, недоліком – неможливість створення *тестових сигналів* бажаного вигляду, більша тривалість експерименту і (іноді) збільшення обсягу обчислень при визначенні параметрів. Наприклад, якщо деякий об'єкт знаходиться під дією 5 незалежних випадкових впливів, кожен з яких розподілений за нормальним законом у діапазоні $[m - 3\sigma, m + 3\sigma]$, а вимірювання параметрів стану здійснюється раз на секунду з похибкою 1%, то для отримання статистично достовірної реакції (тобто розмір вибірки не менше 30) на рідку подію «усі вхідні впливи знаходяться на границі діапазону» необхідно проводити спостереження під час пасивного експерименту безперервно протягом

$$T_{ne} = 30 / \left[\int_{m+3\sigma-0,016\sigma}^{m+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \right]^5 \approx \frac{30}{\left[\frac{0,06}{\sqrt{2\pi}} e^{-\frac{3^2}{2}} \right]^5} =$$

$$= 22565269073747728354 \text{ секунд} \approx 70 \text{ млрд років,}$$

що у 5 разів більше за час існування Всесвіту!

Процес отримання моделі статички у суто математичному сенсі називають апроксимацією. Найпоширенішим методом параметричної ідентифікації є апроксимація за методом найменших квадратів, тобто визначення параметрів евристично заданої функції за критерієм мінімуму середнього квадратичного відхилення цієї функції від експериментальних точок. При дуже маленькій кількості точок (до 5) може також використовуватися інтерполяційний підхід до ідентифікації моделі статички.

Обробка результатів пасивного експерименту при великій кількості даних виконується статистичними методами. Для створення моделей статички при пасивному експерименті використовуються також методи кореляційного і регресійного аналізу, методи оцінювання параметрів моделі на основі критерію максимуму правдоподібності, мінімуму середнього ризику тощо. Ці методи є окремими випадками апроксимації за певних умов: нормальність розподілу ймовірностей експериментальних даних, близькість ідентифікованої залежності до лінійної тощо. Для визначення моделей динаміки об'єкта на основі пасивного експерименту найчастіше використовують зв'язок між спектральною щільністю потужності вхідних і вихідних сигналів і передатною функцією системи. Детальніше методи статистичної ідентифікації розглянуто у розділі 6.

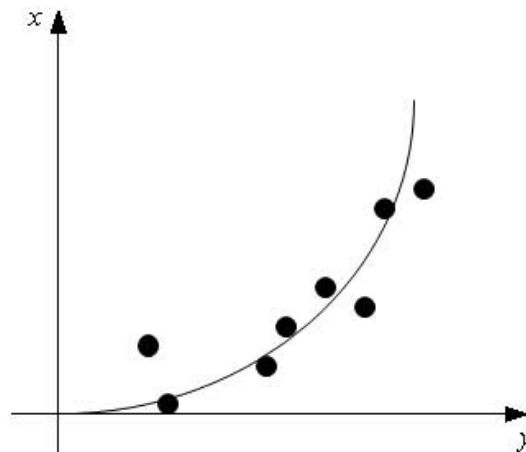
Інтерполяційна ідентифікація найчастіше виконується за допомогою формули Лагранжа

$$y(x) = \sum_{i=0}^n \left[y_i \cdot \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \right] \quad (5.11)$$

де n – кількість експериментальних точок; (x_i, y_i) – значення аргументу і функції в i -тій експериментальній точці.

Очевидно, степінь полінома Лагранжа дорівнює $(n - 1)$. При збільшенні степеня швидко зростає обчислювальна похибка, тому використовувати інтерполяцію для кількості точок $n > 5$ і відповідно, степеня полінома більшого 4 недоцільно.

Приклад апроксимації при невеликій кількості даних (експериментальних точок) показаний на рис. 5.3.



д апроксимації результатів пасивного експерименту при отриманні моделі статики

З рисунка видно, що результати пасивного експерименту розташувалися на площині $y_o(x)$ таким чином, що нагадують параболічну залежність. Таким чином, аналітичний вираз моделі статики будемо шукати у вигляді $y_n = a_2 x^2 + a_1 x + a_0$. Відповідно до методу найменших квадратів критерій пошуку оптимальних коефіцієнтів апроксимації $\{a_0, a_1, a_2\}$ запишеться у вигляді

$$\sigma = \sum_{i=1}^n [a_2 x_i^2 + a_1 x_i + a_0 - y_o(x_i)]^2. \quad (5.12)$$

Після підстановки усіх експериментальних точок отримаємо залежність $\sigma(a_0, a_1, a_2)$. Далі розв'язується задача пошуку таких значень параметрів $\{a_0, a_1, a_2\}$, які забезпечать мінімум СКВ. Для цього можуть застосовуватися різні методи, зокрема аналітичний (на основі прирівнювання до нуля частинних похідних і розв'язування отриманої системи рівнянь), регулярні пошукові методи (наприклад, метод перебору або метод градієнта), методи випадкового пошуку (наприклад, метод Монте-Карло або генетичний алгоритм) та інші.

Детальніше про інтерполяційну та апроксимаційну обробку даних можна прочитати у розділі 7.

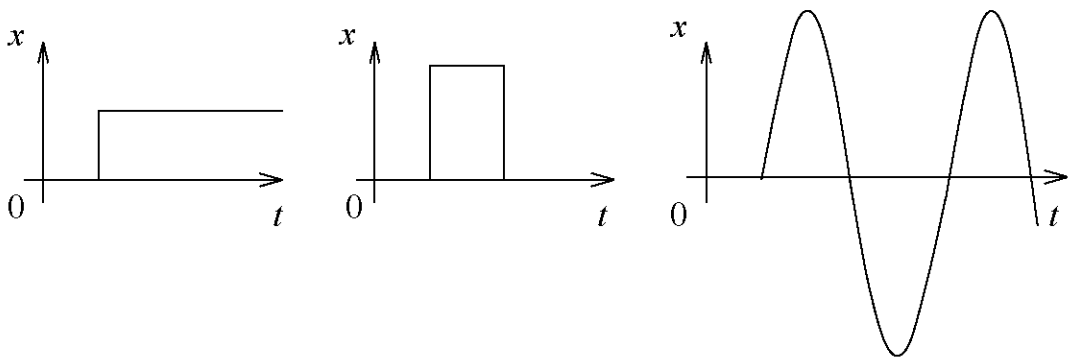
Часто як дані пасивного експерименту використовують записи в експлуатаційних журналах технологічного обладнання. Однак до такої інформації потрібно ставитися критично через можливі помилки в записі, допущені експлуатаційниками, а також похибки за рахунок неодноразової реєстрації даних вимірювань, особливо при великій кількості даних.

5.2.2 Активна ідентифікація функціональної моделі

Схема активного експерименту наведена на рис. 5.2, б.

Як зазначалося вище, активний експеримент передбачає *генерування* діючих на об'єкт *тестових сигналів* потрібної форми, моделювання їх перетворень об'єктом і обробку результатів перетворення. Тестові сигнали поділяються на регулярні і випадкові.

До *регулярних сигналів* відносять *aperiodичні* (ступінчаста функція рис. 5.4, а, прямокутний імпульс рис. 5.4, б та інші) та *періодичної дії* (рис. 5.4, в). Перші використовуються для знаходження коефіцієнтів диференціальних рівнянь, другі – для визначення амплітудно-фазових частотних характеристик об'єктів (рис. 5.4).



При використанні вхідного тестового сигналу у вигляді ступінчастої функції отримують вихідний сигнал, який відтворює перехідну характеристику об'єкта. Залежно від вигляду перехідної характеристики моделі задаються найчастіше одним з трьох видів передавальної 1 функції:

- у вигляді передавальної функції інерційної ланки першого порядку (рис. 5.5, а)

$$W_0(p) = \frac{Ke^{-\tau p}}{Tp + 1}, \quad (5.13)$$

де K, T, τ – коефіцієнт підсилення, постійна часу і запізнювання, що повинні бути визначені в околі номінального режиму роботи об'єкта.

- другого порядку з запізнюванням (рис. 5.5, б)

$$W_0(p) = \frac{Ke^{-\tau p}}{(T_1 p + 1)(T_2 p + 1)}; \quad (5.14)$$

– для об'єкта моделювання без самовирівнювання (рис. 5.5, в) передавальна функція має вигляд

$$W_0(p) = \frac{Ke^{-\tau p}}{p}. \quad (5.15)$$

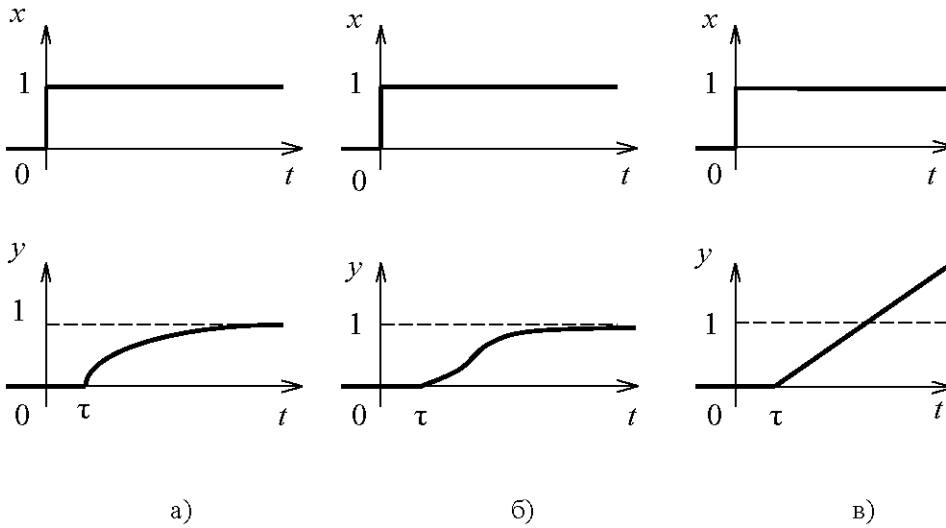
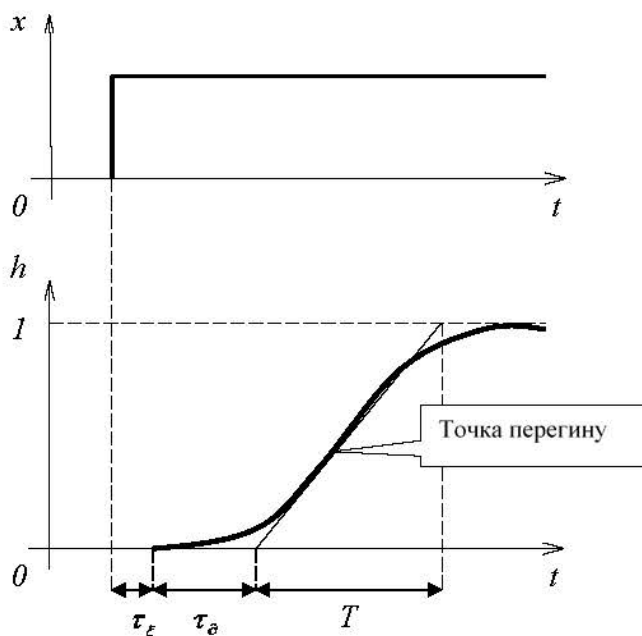


Рисунок 5.5 - Типові перехідні характеристики

Перед початком обробки передатну характеристику рекомендується про- нормувати (діапазон зміни нормованої кривої $[0, 1]$) і виділити з її початкової ділянки величину чистого часового запізнювання. Розглянемо для прикладу нор- мовану перехідну характеристику об'єкта, у якій заздалегідь визначений дина- мічний коефіцієнт підсилення K як відношення збільшення вихідного сигналу до збільшення вхідного в околі робочої точки, після чого перехідну характе- ристику прономровано (рис. 5.6).



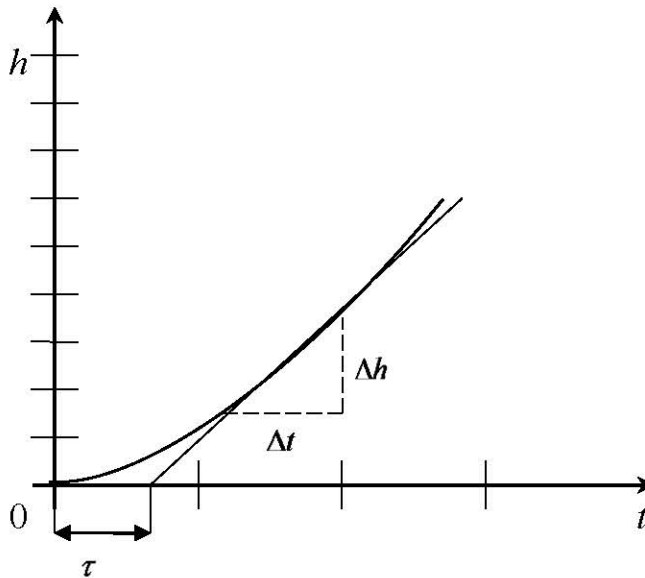
Визначення динамічних характеристик об'єктів за переходною характеристикою можна робити багатьма методами. Застосуємо метод дотичної до точки перегину переходної характеристики.

У даному випадку точка перегину відповідає переходу від режиму прискорення до режиму уповільнення темпу наростання вихідного сигналу. Постійна часу T і динамічне запізнювання τ_δ визначаються за графіком рис. 5.4. Запізнення τ визначаємо як суму чистого запізнювання і динамічного запізнювання, тобто $\tau = \tau_\xi + \tau_\delta$.

Методику визначення параметрів динамічної моделі (5.6) об'єкта без самовирівнювання розглянемо на прикладі рис. 5.7. Для об'єкта без самовирівнювання коефіцієнт підсилення визначається як відношення сталої швидкості зміни вихідної величини до величини стрибка входного сигналу

$$K = \frac{\Delta h / \Delta t}{\Delta x / \Delta t} \quad (5.16)$$

Точнішими, але складнішими методами ідентифікації на основі переходної характеристики є метод Орманса, метод Сімою та інші.



Частотні методи визначення динамічних характеристик ґрунтуються на тому, що на вхід об'єкта подаються періодичні сигнали з відомими частотою й амплітудою. Для цього найчастіше використовуються спеціальні генератори синусоїдальних сигналів з частотою, яка змінюється за лінійним законом, і осцилографи-характеріографи.

Частотні методи визначення динамічних характеристик об'єкта припускають наявність двох етапів, на яких визначаються:

1. Амплітудно-фазова частотна характеристика об'єкта

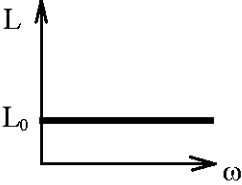
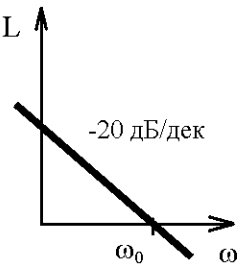
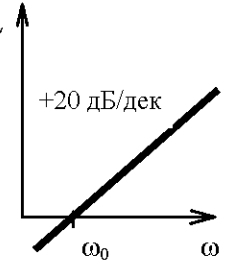
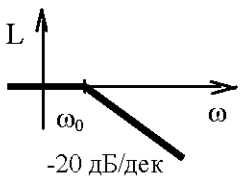
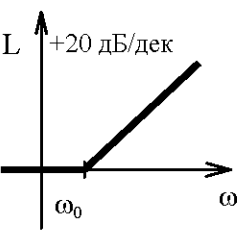
$$A(\omega) = \frac{A_y(\omega)}{A_x(\omega)}, \quad \Phi(\omega) = \varphi_y(\omega) - \varphi_x(\omega), \quad (5.17)$$

де $A_x(\omega)$, $\varphi_x(\omega)$ – амплітуда і фаза вхідного синусоїдального сигналу з кутовою частотою ω ; $A_y(\omega)$, $\varphi_y(\omega)$ – амплітуда і фаза вихідного синусоїдального сигналу;

2. Передавальна функція об'єкта.

Найпростішим способом ідентифікації на основі результатів активного експерименту є побудова асимптотичної логарифмічної амплітудно-частотної характеристики (ЛАЧХ) $L = \lg A(\omega)$ і розбиття її на елементарні складові. Елементарні ланки і відповідні ЛАЧХ і передавальні функції наведені у таблиці 5.2.

Таблиця 5.2 – Типові елементарні ланки

Тип ланки	Передавальна функція	ЛАЧХ	Параметри
Пропорційна	k		$k = e^{L_0}$
Інтегральна	$\frac{1}{Tp}$		$T = \frac{2\pi}{\omega_0}$
Диференціальна	Tp		$T = \frac{2\pi}{\omega_0}$
Аперіодична	$\frac{1}{Tp + 1}$		$T = \frac{2\pi}{\omega_0}$
Форсуюча	$Tp + 1$		$T = \frac{2\pi}{\omega_0}$

Тип ланки	Передавальна функція	ЛАЧХ	Параметри
Коливальна	$\frac{1}{T_1 p^2 + T_2 p + 1}$ $T_2^2 - 4T_1 < 0$		$\omega_0 = \sqrt{ T_2^2 - 4T_1 }$

Загальна передавальна функція отримується як добуток передавальних функцій елементарних ланок.

Приклад знаходження передавальної функції на основі результатів активного експерименту з застосуванням синусоїдальних тестових сигналів змінної частоти показаний на рис. 5.8.

На рис. 5.8, а показана експериментальна амплітудно-частотна характеристика об'єкта моделювання. На рис. 5.8, б показана побудована на її основі асимптотична ЛАЧХ, де $L = \lg A$; $\omega = 2\pi f$, причому кутові частоти відкладені у десятковому логарифмічному масштабі; асимптоти (дотичні) мають нахил, кратний $\pm 20\text{дБ/дек}$; ω_i – частоти перетину асимптот (дотичних). На рис. 5.8, в – показані елементарні складові з таблиці 5.2, на які послідовно розбита асимптотична ЛАЧХ.

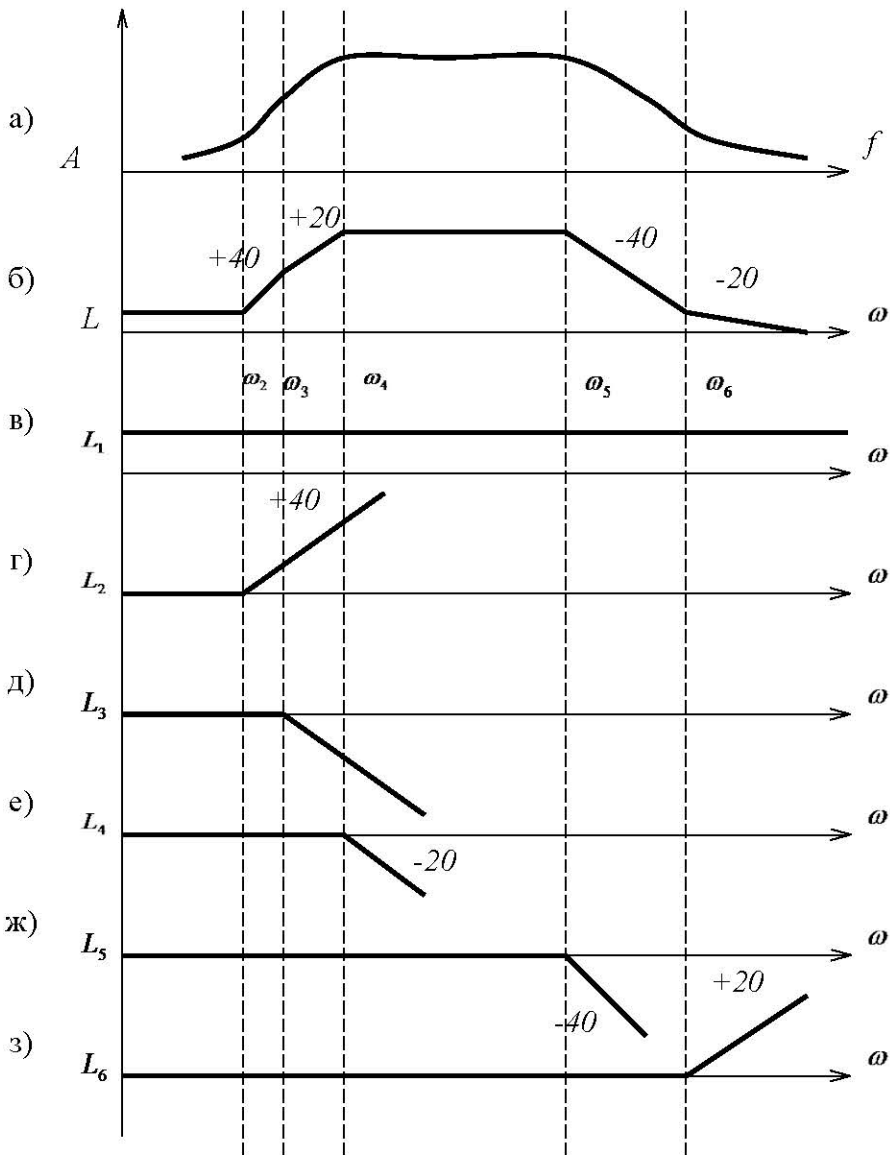
В результаті отримаємо передавальну функцію як добуток передавальних функцій елементарних складових вигляду

$$W(p) = k \cdot \left(\frac{2\pi}{\omega_2} p + 1 \right)^2 \cdot \frac{1}{\frac{2\pi}{\omega_3} p + 1} \cdot \frac{1}{\frac{2\pi}{\omega_4} p + 1} \cdot \left(\frac{1}{\frac{2\pi}{\omega_5} p + 1} \right)^2 \cdot \left(\frac{2\pi}{\omega_6} p + 1 \right).$$

Такий спосіб отримання передавальної функції містить як структурний, так і параметричний етапи ідентифікації, оскільки порядок передавальної функції визначається кількістю і типом елементарних складових, а її параметри – значеннями частот ω_i .

Активний експеримент забезпечує високу достовірність інформації, однак експерименти, пов'язані з дослідженням реакції, наприклад, на ступінчасті або синусоїдальні коливання, важко проводити на типових гідравлічних, енергетичних і теплових об'єктах, оскільки подібні експерименти можуть порушити процес або вивести його зі звичайного режиму в аварійний. Крім того, при про-

веденні активних експериментів важко в реальних виробничих умовах стабілізувати параметри процесу на заданому рівні протягом певного відрізка часу.



5.2.3 Ідентифікація алгоритмічної моделі

Хоча алгоритмічна ідентифікація за змістом близька до структурної ідентифікації, але є набагато складнішою, а у певних випадках ця складність пере-

вищує можливості сучасних комп'ютерів. Прикладом такої задачі є знаходження алгоритму дешифрування інформації при сучасних засобах захисту.

На першому етапі алгоритмічної ідентифікації з'ясовують, чи є алгоритм рекурсивним. Рекурсивний алгоритм відповідає моделі динамічної системи, а нерекурсивний – статичної. Результати роботи нерекурсивного алгоритму залежать лише від поточних вхідних даних. Результати роботи рекурсивного алгоритму залежать як від вхідних даних, так і від попередніх результатів (попередніх вхідних даних). Зауважимо, що рекурсивна модель, в якій поточні результати залежать тільки від безпосередньо попередніх, називається марковською.

У найпростішому випадку нерекурсивного алгоритму ідентифікація здійснюється на основі складання таблиці відповідності «вхідні дані – результати» і обрання методу апроксимації отриманої табличної залежності. Алгоритмічне зображення результату апроксимації і вважають за алгоритмічну модель.

При бінарних (логічних) вхідних і вихідних даних апроксимація звичайно здійснюється за допомогою методу Вейча-Карно. Метод передбачає вибір функціонального базису, складання досконалої диз'юнктивної або кон'юнктивної нормальної форми та її мінімізацію.

Алгоритмічна ідентифікація найчастіше використовується в процесі аналізу алгоритмів, які реалізовані у вигляді програмного забезпечення (програмного коду). Розрізняють динамічний і статичний аналізи.

Динамічний аналіз коду – аналіз програмного забезпечення, що виконується за допомогою виконання програм на реальному або віртуальному процесорі (на відміну від статичного аналізу). Утиліти динамічного аналізу можуть вимагати завантаження спеціальних бібліотек, перекомпіляцію програмного коду. Деякі утиліти можуть документувати виконуваний код в процесі виконання або перед ним. Для більшої ефективності динамічного аналізу потрібне введення до тестованої програми достатньої кількості вхідних даних, щоб дослідити усі розгалуження програмного коду. Також потрібно подбати про мінімізацію впливу документування на виконання програми (враховуючи вплив на часові характеристики).

Приклади утиліт аналізу коду:

- jTracert, Java-агент, який документує код існуючих додатків, що працюють у віртуальній машині JVM, і автоматично будує діаграми послідовності (sequence diagrams);

- Holodeck від компанії Security innovation – інструмент моделювання помилок для динамічного аналізу і тестування уразливості / безпеки програм Windows;

- VB Watch додає до програми код динамічного аналізу в програми мовою Visual Basic для моніторингу їх продуктивності, стека викликів, траси виконання, екземплярів об'єктів, змінних і покриття коду.

Статичний аналіз коду – аналіз програмного забезпечення, який виконується (на відміну від динамічного аналізу) без реального виконання досліджу-

ваних програм. У більшості випадків аналіз проводиться над вихідним кодом, хоча іноді аналізу піддається об'єктний код. Термін зазвичай застосовують до аналізу, який виконується спеціальним програмним забезпеченням, тоді як ручний аналіз називають «program understanding», «program comprehension» (розумінням програми).

Залежно від використовуваного інструмента глибина аналізу може варіюватися від визначення поведінки окремих операторів до аналізу, що містить весь вихідний код.

Останнім часом статичний аналіз все більше використовується у верифікації властивостей програмного забезпечення, використовуваного в комп'ютерних системах високої надійності, особливо критичних для життя (safety-critical). Також застосовується для пошуку коду, що потенційно містить вірусне або інше потенційно небезпечне програмне забезпечення.

5.2.4 Ідентифікація інформаційної моделі

Як зазначалося вище, інформаційна модель – це набори даних (модель статистики) та потоки даних (модель динаміки). Найпоширенішою є ідентифікація структури наборів даних, які подаються у вигляді гіпертексту – структура Інтернет-ресурсів.

Проблема аналізу взаємозв'язків між веб-сайтами є темою досліджень провідних пошукових систем, як Google, Bing, Яндекс тощо. Крім того, низка інших компаній пропонує свої продукти (як комерційні, так і з відкритим програмним кодом) для інтеграції з пошуковими системами для більш ефективного пошуку в мережі.

Сервіс TouchGraph™ GoogleBrowser містить багато корисної інформації (як візуальної, так і статистичної) про веб-сайти та взаємозв'язки між ними. Отримані результати відображаються у вигляді графу. Користувач може заздалегідь задати кількість результатів (до 30) і глибину пошуку зв'язків. Якщо глибина дорівнює 1, то виводиться список сайтів, пов'язаних безпосередньо з вказаним в запиті. Якщо ж глибина пошуку дорівнює 2 і більше (максимальне число рівнів досягає 10), то зв'язки шукаються для першого сайту і для кожного з пов'язаних з ним вузлів другого рівня і т. д.

Сервіс Walk2Web візуалізує взаємозв'язки між сайтами, подаючи процес пошуку інформації. Необхідно тільки ввести в рядок пошуку адресу сайту, що цікавить, і всі посилання, які хоч якось до нього стосуються, будуть у розпорядженні користувача. Після того, як в рядок пошуку була введена назва потрібного сайту, на екрані з'являється невелика схема, за допомогою якої користувач може рухатися далі. Сервіс не працює з пошуковими системами, а показує інформацію лише для заданих сайтів.

Для пошуку зв'язків між змістом веб-сторінок, який не відображений у гіперпосиланнях, використовуються методи семантичного аналізу.

Основою семантики є онтології, що призначені для формалізації предметних областей. Онтологія предметної області описує ключові слова (терміни) предметної області і зв'язок між ними. Аналіз онтологій є складною задачею великої розмірності, оскільки передбачає обробку великих обсягів даних та логічних тверджень, які містяться у онтологіях.

Ідентифікацію потоків даних ще називають аналізом трафіку. Аналізатор трафіку, або сніффер – програмно-апаратний пристрій, призначений для перехоплення і подальшого аналізу, або тільки аналізу мережевого трафіку, призначеного для інших вузлів. Сніффер може аналізувати тільки те, що проходить через його мережеву карту. Також всередині одного сегмента мережі Ethernet всі пакети розсилаються всім машинам, через це можна перехоплювати чужу інформацію.

Перехоплення трафіку може здійснюватися:

- звичайним «прослуховуванням» мережевого інтерфейсу (метод ефективний при використанні в сегменті концентраторів (хабів) замість комутаторів (світців), а також у WiFi-мережах, в іншому випадку метод малоефективний, оскільки на сніффер потрапляють лише окремі пакети);
- підключенням сніффера в розрив каналу;
- відгалуженням (програмним або апаратним) трафіку і спрямуванням його копії на сніффер;
- через аналіз побічних електромагнітних випромінювань і відновлення таким чином трафіку (використовується у засобах електромагнітної розвідки).

Сніфери застосовуються як в позитивних, так і в деструктивних цілях. Аналіз трафіку, який пройшов через сніффер, дозволяє:

- виявити паразитний, вірусний і закільцьований трафік, наявність якого збільшує завантаження мережного обладнання та каналів зв'язку;
- виявити в мережі шкідливе і несанкціоноване програмне забезпечення, наприклад, мережеві сканери, троянські програми тощо;
- перехопити незашифрований (а часом і зашифрований) трафік користувача з метою отримання паролів і іншої інформації;
- локалізувати несправність мережі або помилку конфігурації мережевих агентів.