

УДК 519.872:004.942

**МОДЕЛЮВАННЯ СЕРВЕРНИХ СИСТЕМ ЯК СКЛАДНИХ
МЕРЕЖ МАСОВОГО ОБСЛУГОВУВАННЯ В УМОВАХ
ВИСОКОГО НАВАНТАЖЕННЯ**

Колісник О.В.

e-mail: oleksandr.kolisnyk1@nure.ua

Науковий керівник – канд. техн. наук, доц. Яловега І.Г.

Харківський національний університет радіоелектроніки, каф. ПМ,
м. Харків, Україна

This work investigates the behavior of server systems characterized by non-linear effects that manifest as disproportionate performance degradation under increasing workloads. It is shown that as resource utilization approaches critical thresholds, response times and queue lengths increase sharply. These effects are driven by internal resource constraints. This study formalizes a server system as a network of service nodes where requests transition between stages, forming queues. Ultimately, the server system is represented as a dynamic object whose state evolves over time under the influence of input request flows and internal processing mechanisms.

Серверні системи характеризуються наявністю нелінійних ефектів, що проявляються у непропорційній зміні продуктивності при зростанні навантаження. Зокрема, при наближенні до граничних значень завантаження ресурсів спостерігається різке зростання часу відгуку та довжини черг, що не може бути описано лінійними моделями. Така поведінка зумовлена як внутрішніми обмеженнями ресурсів, так і складними механізмами взаємодії між компонентами системи, включаючи конкуренцію за доступ до пам'яті, блокування даних та синхронізацію процесів [1, 2].

У межах системного підходу Дж. Клейнрока серверна система розглядається як сукупність обслуговуючих серверів, між якими переміщуються заявки (запити), утворюючи черги та проходячи послідовні етапи обробки [3]. Така формалізація дозволяє перейти від якісного опису функціонування системи до кількісного аналізу її характеристик, зокрема визначення пропускної здатності, часу очікування та ймовірності відмови в обслуговуванні.

Поєднання підходів загальної теорії систем і теорії масового обслуговування створює методологічну основу для дослідження серверних систем як складних динамічних об'єктів, що функціонують у умовах невизначеності та змінного навантаження. Це дозволяє не лише адекватно описати їхню поведінку, але й сформувати передумови для розробки математичних моделей, спрямованих на аналіз критичних режимів функціонування, зокрема процесів насичення ресурсів, втрати стійкості та деградації продуктивності.

Отже, серверну систему можна представити як динамічний об'єкт, стан якого змінюється у часі під впливом вхідного потоку запитів та внутрішніх механізмів обробки. Формально стан системи доцільно описати вектором:

$$X(t) = \{Q(t), \lambda(t), \mu(t), R(t), L(t)\},$$

де $Q(t)$ – довжина черги запитів; $\lambda(t)$ – інтенсивність вхідного потоку; $\mu(t)$ – інтенсивність обслуговування; $R(t)$ – вектор доступних ресурсів; $L(t)$ – стан механізмів синхронізації (блокування, latch тощо).

Для представлення процесів обробки запитів у серверній системі доцільно подати її у вигляді багатоланкової мережі масового обслуговування, що відображає послідовність проходження запитів через основні функціональні вузли системи та їх взаємозв'язки. Узагальнену структуру серверної системи як мережі масового обслуговування наведено на рис. 1.

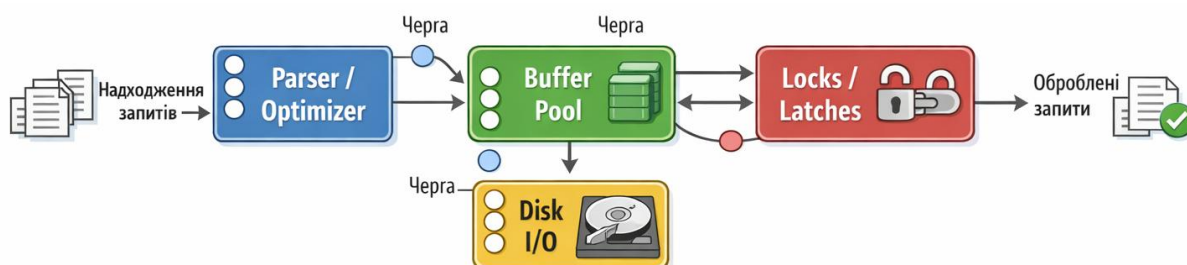


Рисунок 1 – Схема серверної системи як мережі масового обслуговування

Як видно з рис. 1, процес обробки запитів у серверній системі має послідовно-паралельний характер і реалізується через сукупність взаємопов'язаних вузлів обслуговування. Запити надходять до системи у вигляді вхідного потоку та спочатку обробляються у вузлі синтаксичного та семантичного аналізу (parser/optimizer), після чого передаються до підсистеми кешування. У разі відсутності необхідних даних у буферній пам'яті відбувається звернення до дискової підсистеми.

Список використаних джерел:

1. Bertalanffy L. von. General System Theory: Foundations, Development, Applications. New York : George Braziller, 1968. 289 p.
2. Tanenbaum A. S., Bos H. Modern Operating Systems. 4th ed. Upper Saddle River : Pearson, 2015. 1136 p.
3. Kleinrock L. Queueing Systems. Volume 1: Theory. New York: Wiley-Interscience, 1975. 417 p.