

Аспірант Пасічник М.Ю.; д.т.н., проф. Зайцев В.Г.

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»

## ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ ДЛЯ ДИСПЕТЧЕРИЗАЦІЇ ЗАВДАНЬ У СИСТЕМАХ РЕАЛЬНОГО ЧАСУ

### Abstract

*M.Y. Pasichnyk; V.H. Zaitsev*

#### ***Load Prediction for Task Scheduling in Real-Time Systems***

*The paper presents an approach to integrating short-term load prediction into real-time task scheduling. The proposed concept aims to improve responsiveness to dynamic workloads while preserving timing guarantees for periodic tasks. By adapting scheduling parameters based on forecasted demand, the system can maintain stability under variable conditions, reduce delays for event-driven requests, and achieve a more efficient balance between resource utilization and timeliness.*

### Вступ

Системи реального часу одночасно обслуговують синхронні завдання з жорсткими дедлайнами та асинхронні завдання зі змінним часом появи. Статично налаштовані системи реального часу в межах або призводять до зайвих резервів, або до великих затримок під час короткочасних сплесків надходження завдань. У критичних застосуваннях (керування процесами, транспорт, енергетика, медичні пристрої) потрібна система, що зменшує затримку виконання асинхронних завдань.

Проблема поєднання прогнозування та диспетчеризації завдань у системах реального часу перебуває в центрі уваги сучасних досліджень. У низці робіт розглядаються різні аспекти використання прогностичної інформації для поліпшення ефективності планування. У роботі [1] зроблено акцент на підвищенні стійкості алгоритмів планування тривалості виконання із прогнозами. Автори вказують, що навіть за наявності якісних прогнозів необхідні додаткові механізми для компенсації похибок, аби уникати зривів дедлайнів. Цей підхід важливий у контексті систем реального часу, де надійність є критичною. У роботі [2] розглядається можливість інтеграції прогнозів у політики диспетчеризації реального часу. Автори пропонують підходи, що дозволяють коригувати параметри планування на основі короткострокових оцінок навантаження,

при цьому зберігаючи гарантію передбачуваності для періодичних завдань. Особливу увагу приділено питанням безпечності адаптації параметрів. У роботі [3] аналізується розклад на однорідних машинах із використанням прогнозів. Доведено, що навіть неточні прогнози можуть суттєво знизити середній час завершення завдань за умови правильного обліку похибки. Такий підхід є релевантним для сценаріїв, де аперіодичні завдання складають значну частку навантаження. У публікації [4] досліджуються алгоритми планування, де система не має повної інформації про майбутнє, але використовує наближені оцінки для прогнозування навантаження. Це дозволяє скорочувати затримки в середньому випадку, водночас зберігаючи прийнятну надійність у гірших сценаріях. Таким чином, сучасні дослідження сходяться в тому, що прогнози можуть істотно зменшити затримки і підвищити пропускну здатність, але лише за умови обережного застосування та наявності механізмів, що компенсують похибки.

У роботі було запропоновано використовувати прогнозування навантаження для підвищення ефективності диспетчеризації в системах реального часу. Ідея полягає в тому, щоб за допомогою короткострокових прогнозів вчасно виявляти сплески появи завдань і відповідно налаштовувати політику обслуговування так, аби зберігалися гарантії для регулярних завдань і водночас скорочувались затримки для асинхронних завдань.

### **Постановка завдання**

Ефективна диспетчеризація з урахуванням короткострокового навантаження має поєднувати коригування параметрів обслуговування з незмінними гарантіями своєчасності. Мета — використовувати короткострокову оцінку навантаження як допоміжний сигнал для прийняття рішень і застосовувати зміни лише після підтвердженої придатності розкладання для обраної політики. Основні вимоги та виклики включають:

- Узгодженість із гарантіями.
- Метрика та часовий інтервал.
- Запас на похибку.
- Правила адаптації.
- Режим при невизначеності.
- Накладні витрати та пріоритетність.
- Оцінка результатів і межі застосовності.

### Виклад основного матеріалу

Основна ідея полягає у введенні короткострокового прогнозу навантаження на проміжку .

Нехай  $A(t, \Delta t)$  — множина аперіодичних запитів, що надійшли у вікні . Для оцінки запропоновано використовувати сумарний попит на процесорний час

$$\hat{C}(t, \Delta t) = \sum_{i \in A(t, \Delta t)} C_i$$

де  $C_i$  — оцінка часу виконання запиту  $i$  (використовується останнє виміряне значення або ковзне усереднення).

Щоб компенсувати невизначеність прогнозу, вводимо запас на похибку  $\delta$ , тоді оцінка набуває вигляду:

$$\hat{C}^*(t, \Delta t) = \hat{C}(t, \Delta t) + \delta$$

Інтеграція з планувальником відбувається через порівняння  $\hat{C}^*(t, \Delta t)$  з доступним у цьому ж вікні бюджетом для аперіодики  $B(t, \Delta t)$ , який впливає з обраної базової політики планування. Визначаємо індекс навантаження, що виступає керуючим сигналом:

$$\rho(t, \Delta t) = \frac{\hat{C}^*(t, \Delta t)}{B(t, \Delta t)}$$

Щоб система не перемикалася між режимами надто часто, вводимо гістерезис із двома порогами  $\tau_-$  і  $\tau_+$  та обмежуємо швидкість зміни  $B$ , що стабілізує реакцію системи й підтримує передбачуваність:

- Якщо  $\rho < \tau_-$  система працює у звичному режимі.
- Якщо  $\tau_- \leq \rho \leq \tau_+$  доцільне тимчасове розширення бюджету для аперіодичних запитів  $B$  за умови збереження здійсненності.
- Якщо  $\rho > \tau_+$ , застосовуються м'які обмеження аперіодики (відтермінування низькопріоритетних запитів, зменшення квот).

Таким чином, прогнозування навантаження у короткому інтервалі часу можна розглядати як своєрідний механізм раннього попередження. Воно не змінює фундаментальних гарантій системи реального часу, але надає можливість покращити середні показники якості обслуговування, зменшити затримки для аперіодичних завдань і підвищити стійкість системи до змінних умов.

## Висновки

Використання прогнозування навантаження може підвищити ефективність роботи системи без порушення основних гарантій своєчасності. Основні принципи полягають у перевірці придатності перед будь-яким коригуванням, у застосуванні запасу на похибку та в чіткому визначенні параметрів, які допускають зміну. Такий підхід забезпечує зменшення затримок для аперіодичних завдань, збереження пріоритету періодичних завдань і стабільність роботи навіть за умов раптових змін навантаження. Подальші дослідження варто спрямувати на розробку методів точнішого прогнозування, оцінку накладних витрат та визначення меж застосовності підходу в різних класах систем.

## Література

1. Zhao T., Li C., Li W., Zomaya A. Y. Brief Announcement: Towards a More Robust Algorithm for Flow Time Scheduling with Predictions. Proceedings of the 34th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA). – July 2022. – DOI: 10.1145/3490148.3538557.
2. Zhao T., Li W., Zomaya A. Y. Real-Time Scheduling with Predictions. Proceedings of the 2022 IEEE Real-Time Systems Symposium (RTSS). – December 2022. – DOI: 10.1109/RTSS55097.2022.00036.
3. Zhao T., Li W., Zomaya A. Y. Uniform Machine Scheduling with Predictions. Proceedings of the Thirty-Second International Conference on Automated Planning and Scheduling (ICAPS). – July 2022.
4. Im S., Kumar R., Montazer Qaem M., Purohit M. Non-Clairvoyant Scheduling with Predictions. Proceedings of the 33rd ACM Symposium on Parallelism in Algorithms and Architectures (SPAA). – July 2021. – DOI: 10.1145/3409964.3461790.