

ABSTRACT

Explanatory note size – 108 pages, contains 17 illustrations, 10 tables, 5 applications.

Topicality. In the modern information space, an increasingly important role is played by real-time communication systems that provide instant data exchange between users. The growth in data volumes and the number of users leads to higher requirements for the performance, scalability, and reliability of such systems. Traditional approaches based on blocking solutions are not always able to ensure the required level of performance under high load. The use of a reactive technology stack can significantly improve these parameters, reducing resource consumption for processing a large number of parallel connections.

The aim of the study. Increasing the performance, scalability, and fault tolerance of real-time communication systems that must process and maintain a large number of concurrent connections.

Object of research: the architecture of high-load real-time communication systems.

Subject of research: software architectures and methods of increasing the performance and scalability of high-load real-time communication systems.

To achieve this goal, the **following tasks** were formulated:

- analysis of existing solutions;
- formulation of requirements and criteria for the architecture of a real-time communication system, taking into account reactive principles and modern scalability challenges;
- development of an appropriate architecture and an algorithm for balancing persistent connections based on domain-specific metrics with support for safe, lossless session recovery;
- modeling and experimental evaluation of the effectiveness of the proposed solution in comparison with traditional approaches and existing systems.

The scientific novelty. A new architecture of high-load real-time communication systems is proposed through the use of a reactive technology stack and improved load-balancing methods to optimize the processing and maintenance of a large number of parallel connections.

The practical value. The architecture proposed as a result of the study can be used by developers to build efficient and scalable real-time communication systems. This will help improve the quality of services that require processing a large number of simultaneous connections, increase server fault tolerance under growing load, and reduce server resource costs.

Relationship with working with scientific programs, plans, topics. Work was performed at the Department of Informatics and Software Engineering of the National Technical University of Ukraine «Kyiv Polytechnic Institute. Igor Sikorsky».

Approbation. The scientific provisions of the dissertation were tested at the First All-Ukrainian Scientific and Practical Conference of Young Scientists and Students "Software Engineering and Advanced Information Technologies" (SoftTech-2025) - Kyiv, Ukraine.

Publications. The scientific provisions of the dissertation published in:

– Hnitiy Ye.O., Kramar Yu.M. Architecture of high-load reactive real-time communication systems. Proceedings of the 9th International Scientific and Practical Conference of Young Scientists and Students “Software Engineering and Advanced Information Technologies (SoftTech-2025)”, Section of the Department of Informatics and Software Engineering, 26–28 November 2025, Kyiv.

Keywords: ARCHITECTURE, LOAD BALANCING, GRACEFUL RECONNECT, KUBERNETES, WEBSOCKET, SCALING