ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ
INTERNATIONAL UNIVERSITY
SINCE 2009
МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (21) 1
ақпан - наурыз

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

БАС РЕДАКТОР:

Исахов Асылбек Абдиашимович — басқарма төрағасы, Халықаралық ақпараттық технологиялар универси тетінің ректоры, есептеу теориясы саласындағы математика бойынша PhD докторы, “Компьютерлік ғылымдар және информатика”бағыты бойынша қауымдастырылған профессор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті, «Ақпараттық жүйелер» кафедрасының проректоры (Қазақстан)

ҒАЛЫМ ХАТШЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, «Халықаралық ақпараттық технологиялар университеті» АҚ, Ғылыми-зерттеу жұмыс департаментінің директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

Разак Абдул — PhD, Халықаралық ақпараттық технологиялар университетінің профессоры (Қазақстан)

Лучио Томмазо де Паолис — Саленто университетінің (Италия) инновациялар және технологиялық инженерия департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры

Лиз Бэкон — профессор, Абертей университеті вице-канцлердің орынбасары (Ұлыбритания)

Микеле Пагано — PhD, Пиза университетінің профессоры (Италия)

Отелбаев Мухтарбай Отелбаевич — физика-математика ғылымдарының докторы, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті, «Математикалық және компьютерлік модельдеу» кафедрасының профессоры (Қазақстан)

Рысбайулы Болатбек — физика-математика ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті, «Математикалық және компьютерлік модельдеу» кафедрасының профессоры (Қазақстан)

Дайнеко Евгения Александровна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің Жаһандық серіктестік және қосымша білім беру жөніндегі проректоры (Қазақстан)

Дузбаев Нуржан Токкужаевич — PhD, Халықаралық ақпараттық технологиялар университетінің Цифрландыру және инновациялар жөніндегі проректоры (Қазақстан)

Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының профессоры (Қазақстан)

Сейлова Нұргүл Абдуллаевна — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Компьютерлік технологиялар және киберқауіпсіздік» факультетінің деканы (Қазақстан)

Мухамедиева Ардақ Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Цифрлық трансформациялар» факультетінің деканы (Қазақстан)

Ыдырыс Айжан Жұмабайқызы — PhD, Халықаралық ақпараттық технологиялар университетінің «Математикалық және компьютерлік модельдеу» кафедрасының меңгерушісі (Қазақстан)

Шильдибеков Ерлан Жаржанович — PhD, Халықаралық ақпараттық технологиялар университетінің «Экономика және бизнес» кафедрасының меңгерушісі (Қазақстан)

Аманжолова Сауле Токсановна — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Киберқауіпсіздік» кафедрасының меңгерушісі (Қазақстан)

Ниязгулова Айгүл Аскарбековна — филология ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Медиакоммуникациялар және Қазақстан тарихы» кафедрасының меңгерушісі (Қазақстан)

Айтмагамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университетінің «Радиотехника, электроника және телекоммуникация» кафедрасының профессоры (Қазақстан)

Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университетінің қауымдастырылған профессоры (Қазақстан)

Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры (Қазақстан)

Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)

Тадеуш Валлас — PhD, Адам Мицкевич атындағы университеттің проректоры (Польша)

Мамырбаев Өркен Жұмажанұлы — Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялары институты директорының ғылым жөніндегі орынбасары (Қазақстан)

Бушуев Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның «УКРНЕТ» жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сәулет университетінің «Жобаларды басқару» кафедрасының меңгерушісі (Украина)

Белощицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Астана IT университетінің деректер жөніндегі есептеу және ғылым кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Мрзабаева Раушан Жәлиқызы — «Халықаралық ақпараттық технологиялар университеті» АҚ (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншіктенуші: «Халықаралық ақпараттық технологиялар университеті» АҚ (Алматы қ.).
Қазақстан Республикасы Ақпарат және әлеуметтік даму министрлігінің Ақпарат комитетінде – **20.02.2020** жылы берілген.
№ KZ82VPY00020475 мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: ақпараттық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологияларға арналған.
Мерзімділігі: жылына 4 рет.
Тиражы: 100 дана
Редакцияның мекенжайы: 050040, Алматы қ-сы, Манас к-сі, 34/1, 709-кабинет, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Журнал сайты: https://journal.iitu.edu.kz

© Халықаралық ақпараттық технологиялар университеті АҚ, 2025
© Авторлар ұжымы, 2025

ГЛАВНЫЙ РЕДАКТОР:

Исахов Асылбек Абдиашимович — кдоктор PhD по математике в области теории вычислимости, ассоциированный профессор по направлению "Компьютерные науки и информатика", Председатель Правления – Ректор АО «Международный университет информационных технологий» (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследо ва тель-ской деятельности Международного университета информационных технологий (Казахстан)

УЧЕНЫЙ СЕКРЕТАРЬ:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департа мента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных тех нологий (Казахстан)

Лучио Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайулы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайулы Болатбек — доктор физико-математических наук, профессор, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Дайнеко Евгения Александровна — PhD, ассоциированный профессор, проректор по глобальному партнерству и дополнительному образованию Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета цифровых трансформаций Международного университета информационных технологий (Казахстан)

Ыдырыс Айжан Жумабаевна — PhD, ассистент профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шилдибеков Ерлан Жаржанович — PhD, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Аманжолова Сауле Токсановна — кандидат технических наук, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, заведующая кафедрой медиакоммуникаций и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Зуфарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Алмисреб Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Тадеуш Валлас – PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белощицкая Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — АО «Международный университет информационных технологий» (Казахстан).

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (г. Алматы).
Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**
Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**
Периодичность: 4 раза в год.
Тираж: 100 экземпляров.
Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz
© АО Международный университет информационных технологий, 2025
© Коллектив авторов, 2025

EDITOR-IN-CHIEF:

Isakhov Asylbek Abdiashimovich — PhD in Mathematics specializing in Computability Theory and Associate Professor in Computer Science and Informatics, Chairman of the Board, Rector of International Information Technology University (Kazakhstan)

DEPUTY CHIEF DIRECTOR:

Kolesnikova Katerina Viktorovna — Doctor of Technical Sciences, Vice-Rector of Information Systems Department, International Information Technology University (Kazakhstan)

SCIENTIFIC SECRETARY:

Ipalakova Madina Tulegenovna — Candidate of Technical Sciences, Associate Professor, Director of the Research Department, International University of Information Technologies (Kazakhstan)

EDITORIAL BOARD:

Razaq Abdul — PhD, Professor of International Information Technology University (Kazakhstan)

Lucio Tommaso de Paolis — Director of Research and Development, AVR Laboratory, Department of Innovation and Process Engineering, University of Salento (Italy)

Liz Bacon — Professor, Deputy Director, and Deputy Vice-Chancellor of the University of Abertay. (Great Britain)

Michele Pagano — Ph.D., Professor, University of Pisa (Italy)

Otelbaev Mukhtarbay Otelbayuly – Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of the Republic of Kazakhstan, Professor of the Department of Mathematical and Computer Modeling of International Information Technology University (Kazakhstan)

Rysbayuly Bolatbek — Doctor of Physical and Mathematical Sciences, Professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Daineko Yevgeniya Alexandrovna — PhD, Associate Professor, Vice-Rector for Global Partnership and Continuing Education, International Information Technology University (Kazakhstan)

Duzbaev Nurzhan Tokkuzhaevich — Candidate of Technical Sciences, Vice-Rector for Digitalization and Innovations, International Information Technology University (Kazakhstan)

Sinchev Bakhtgerey Kuspanuly — Doctor of Technical Sciences, Professor of the Department of Information Systems, International Information Technology University (Kazakhstan)

Seilova Nurgul Abdullaevna — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)

Mukhamedieva Ardak Gabitovna – Candidate of Economic Sciences, Dean of the Faculty of Digital Transformations, International Information Technology University (Kazakhstan)

Idyrys Aizhan Zhumabaevna — PhD, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)

Shildibekov Yerlan Zharzhanuly — PhD, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)

Amanzholova Saule Toksanovna — Candidate of Technical Sciences, Head of the Department of Cyber Security, International Information Technology University (Kazakhstan)

Niyazgulova Aigul Askarbekovna — Candidate of Philology, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)

Aitmagambetov Altai Zufarovich — Candidate of Technical Sciences, Professor of the Department of Radioengineering, Electronics and Telecommunication, International Information Technology University (Kazakhstan)

Almisreb Ali Abd — PhD, Associate Professor, International Information Technology University (Kazakhstan)

Mohamed Ahmed Hamada — PhD, Associate Professor, Department of Information systems, International Information Technology University (Kazakhstan)

Young Im Choo — PhD, Professor, Gachon University (South Korea)

Tadeusz Wallas — PhD, University of Dr. Litt Adam Miskevich in Poznan (Poland)

Mamyrbayev Orken Zhumazhanovich — PhD in Information Systems, Deputy Director for Science, Institute of Information and Computing Technologies CS MSHE RK (Kazakhstan)

Bushuyev Sergey Dmitriyevich — Doctor of Technical Sciences, Professor, Director of Uдоктор технических наук, профессор, директор Ukrainian Association of Project Management UKRNET, Head of Project Management Department, Kyiv National University of Construction and Architecture (Ukraine)

Beloshitskaya Svetlana Vasilyevna — Doctor of Technical Sciences, Associate Professor, Professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)

EXECUTIVE EDITOR

Mrzabayeva Raushan Zhalievna — International Information Technology University (Kazakhstan)

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

© International Information Technology University JSC, 2025
© Group of authors, 2025

МАЗМҰНЫ

ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

СОДЕРЖАНИЕ

ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИ СОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

CONTENT

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

INFORMATION TECHNOLOGY

INFORMATION SECURITY AND COMMUNICATION TECHNOLOGIES

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 1. Number 21 (2025). Pp. 185–200
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.21.1.013

УДК 004.89

FAULT TOLERANCE AND RELIABILITY IN KUBERNETES-ORCHESTRATED MULTI-AGENT SYSTEMS: UNIVERSITY SCHEDULING CASE STUDY

B. Kumalakov, A. Kaziz*
Astana IT University, Astana, Kazakhstan.
E-mail: 231796@astanait.edu.kz

Bolatzhan Kumalakov — PhD in Computer Science, associate professor, the Department of Computational and Data Sciences, Astana IT University, Astana, Kazakhstan.
ORCID: 0000-0003-1476-9542;
Alisher Kaziz — master's student, Department of Computer Engineering, Astana IT University, Astana, Kazakhstan.
E-mail: 231796@astanait.edu.kz. ORCID: 0009-0003-2554-6149.

Abstract. Multi-Agent Systems play a particular role in distributed computing and in environments requiring autonomous coordination, such as robotics, cloud computing, and traffic management. However, ensuring fault tolerance and reliability in MAS remains a significant challenge, particularly in large-scale deployments. This study investigates the impact of Kubernetes-based orchestration on the fault tolerance of MAS, evaluating mechanisms such as automated scaling, redundancy strategies, and self-healing capabilities. Experimental results demonstrate that Kubernetes enhances MAS resilience by reducing failure frequency and improving Mean Time to Recovery. The study also identifies trade-offs between performance and resource consumption, showing that while redundancy and auto-scaling improve system robustness, they introduce computational overhead. Affinity-based scheduling and selective redundancy strategies were found to balance efficiency and reliability effectively. The findings have significant implications for real-world MAS deployments, particularly in optimizing Kubernetes configurations to achieve fault tolerance without excessive resource utilization. Future research should focus on AI-driven scaling, hybrid cloud-edge execution, and enhanced fault detection mechanisms to further improve MAS reliability and efficiency in dynamic environments.

Keywords: machine learning, MAS, MAS Optimization, fault detection, MAS maintenance, cloud-native deployment

For citation: B. Kumalakov, A. Kaziz. FAULT TOLERANCE AND RE-

LIABILITY IN KUBERNETES-ORCHESTRATED MULTI-AGENT SYSTEMS: UNIVERSITY SCHEDULING CASE STUDY//INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES. 2025. Vol. 6. No. 21. Pp. 185–200 (In Eng.). https://doi.org/10.54309/IJICT.2025.21.1.013.

ҒИМАРАТТАРДАҒЫ KUBERNETES АРҚЫЛЫ ОРКЕСТРЛЕНГЕН КӨПАГЕНТТІК ЖҮЙЕЛЕРДЕГІ АҚАУҒА ТӨЗІМДІЛІК ПЕН СЕНІМДІЛІК: УНИВЕРСИТЕТТІҢ КЕСТЕСІН ЖОСПАРЛАУ КЕЙСІ

Б.А. Кумалаков, А.Б. Казиз*
Astana IT University, Астана, Қазақстан.
E-mail: 231796@astanait.edu.kz

Кумалаков Болатжан Арменұлы — PhD Информатика, «Есептеу және деректер ғылымы» департаментінің қауымдастырылған профессор, Astana IT University
ORCID: 0000-0003-1476-9542;
Қазиз Әлішер Бекболатұлы— Магистрант, «**К**омпьютерлік инженерия» департаменті, Astana IT University
E-mail: 231796@astanait.edu.kz. ORCID: 0009-0003-2554-6149.

Аннотация. Көпагентті жүйелер (MAS) таратылған есептеулерде және автономды үйлестіруді қажет ететін ортада, мысалы, робототехника, бұлттық есептеулер және көлік қозғалысын басқаруда ерекше рөл атқарады. Алайда, MAS-тың ақауға төзімділігі мен сенімділігін қамтамасыз ету, әсіресе ауқымды енгізулерде, маңызды мәселе болып қала береді. Бұл зерттеуде Kubernetes негізіндегі оркестрацияның MAS-тың ақауға төзімділігіне әсері зерттеліп, автоматты масштабтау, артықтық стратегиялары және өздігінен қалпына келтіру мүмкіндіктері сияқты механизмдер бағаланады. Эксперименттік нәтижелер көрсеткендей, Kubernetes ақау жиілігін азайтып, қалпына келтірудің орташа уақытын жақсарту арқылы MAS-тың тұрақтылығын арттырады. Зерттеу барысында өнімділік пен ресурстарды тұтыну арасындағы айырбастар да анықталды, өйткені артықтық және автоматты масштабтау жүйенің сенімділігін арттырғанымен, есептеу жүктемесін көбейтеді. Аффинді жоспарлау мен таңдаулы артықтық стратегиялары тиімділік пен сенімділіктің тепе-теңдігін сақтауға көмектесетіні анықталды. Бұл зерттеу нәтижелері нақты MAS енгізулеріне үлкен әсер етеді, әсіресе Kubernetes конфигурацияларын оңтайландырып, артық ресурстарды пайдаланбай ақауға төзімділікті қамтамасыз етуге бағытталған. Болашақ зерттеулер ИИ басқаратын масштабтауға, гибридті бұлт-перифериялық орындауға және ақауларды анықтаудың жетілдірілген механизмдеріне бағытталуы тиіс, осылайша MAS-тың сенімділігі мен тиімділігін динамикалық ортада жақсарту көзделеді.

Түйін сөздер: Машиналық оқыту, Көпагенттік жүйе, Көпагенттік жүйелерді оңтайландыру, Ақауларды анықтау, Көпагенттік жүйелерді қызмет көрсету, Бұлтты орналастыру

Дәйексөздер үшін: К.Б. Арменұлы, Қ.Ә. Бекболатұлы. ҒИМАРАТТАРДАҒЫ KUBERNETES АРҚЫЛЫ ОРКЕСТРЛЕНГЕН КӨПАГЕНТТІК ЖҮЙЕЛЕРДЕГІ АҚАУҒА ТӨЗІМДІЛІК ПЕН СЕНІМДІЛІК: УНИВЕРСИТЕТТІҢ КЕСТЕСІН ЖОСПАРЛАУ КЕЙСІ//ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ. 2025. Т. 6. No. 21. 185–200 бет. (ағылшын тілінде). https://doi.org/10.54309/IJICT.2025.21.1.013.

ОТКАЗОУСТОЙЧИВОСТЬ И НАДЕЖНОСТЬ В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ, ОРКЕСТРИРУЕМЫХ KUBERNETES: КЕЙС РАСПИСАНИЯ УНИВЕРСИТЕТА

Б.А. Кумалаков, А.Б. Казиз*
Astana IT University, Астана, Казахстан.
E-mail: 231796@astanait.edu.kz

Кумалаков Болатжан Арменулы — PhD, ассоциированный профессор департамента «Вычислений и науки о данных», Astana IT University , Астана, Казахстан.
ORCID: 0000-0003-1476-9542;
Қазиз Әлішер Бекболатұлы — магистрант, департамент «Компьютерной инженерии», Astana IT University, Астана, Казахстан.
E-mail: 231796@astanait.edu.kz. ORCID: 0009-0003-2554-6149.

Аннотация. Многоагентные системы (MAS) играют особую роль в распределенных вычислениях и в средах, требующих автономной координации, таких как робототехника, облачные вычисления и управление дорожным движением. Однако обеспечение отказоустойчивости и надежности MAS остается значительной задачей, особенно в крупномасштабных развертываниях. В данном исследовании рассматривается влияние оркестрации на основе Kubernetes на отказоустойчивость MAS, оцениваются такие механизмы, как автоматическое масштабирование, стратегии избыточности и самовосстановление. Экспериментальные результаты показывают, что Kubernetes повышает устойчивость MAS, снижая частоту сбоев и сокращая среднее время восстановления. В исследовании также рассматриваются компромиссы между производительностью и потреблением ресурсов, показывая, что избыточность и авто-масштабирование улучшают надежность системы, но при этом увеличивают вычислительную нагрузку. Планирование на основе аффинности и выборочная избыточность были признаны эффективными методами балансировки эффективности и надежности. Полученные результаты имеют важное значение для реальных внедрений

MAS, особенно в оптимизации конфигураций Kubernetes для обеспечения отказоустойчивости без чрезмерного расхода ресурсов. Будущие исследования должны сосредоточиться на масштабировании, управляемом ИИ, гибридном облачно-периферийном исполнении и усовершенствованных механизмах обнаружения сбоев для дальнейшего повышения надежности и эффективности MAS в динамичных средах.

Ключевые слова: машинное обучение, мультиагентная система, оптимизация мультиагентных систем, обнаружение неисправностей, обслуживание мультиагентных систем, облачное развертывание.

Для цитирования: Б.А. Кумалаков, А.Б. Казиз ОТКАЗОУСТОЙЧИВОСТЬ И НАДЕЖНОСТЬ В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ, ОРКЕСТРИРУЕМЫХ KUBERNETES: КЕЙС РАСПИСАНИЯ УНИВЕРСИТЕТА//МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ. 2025. Т. 6. No. 21. Стр. 185–200. (На англ.). https://doi.org/10.54309/IJICT.2025.21.1.013.

Introduction

Multi-Agent Systems (MAS) have gained significant attention in distributed computing due to their ability to coordinate multiple intelligent agents in dynamic environments (Tolstosheyev et al., 2024). These systems are widely applied in various domains, including robotics, traffic management, and cloud computing (Collier et al., 2019; Heintzman, 2022). However, as MAS become more complex and deployed in large-scale environments, ensuring fault tolerance and reliability becomes a critical challenge. The advent of cloud-native technologies, particularly Kubernetes, has provided new avenues to enhance the resilience of MAS by leveraging container orchestration for automated scaling, self-healing, and load balancing (Dähling, Razik & Monti, 2021; Gu et al., 2025).

Fault tolerance in MAS refers to the system's ability to continue operating correctly despite failures in individual agents, communication networks, or computing infrastructure. Failures in MAS can arise due to network disruptions, software bugs, hardware malfunctions, or resource constraints. Kubernetes, as a container orchestration platform, offers built-in mechanisms such as automated restarts, health monitoring, and redundant deployments that can mitigate such failures (Casquero et al., 2019; Senjab et al., 2023). However, while Kubernetes enhances system reliability, the complexity of managing MAS within this environment introduces new challenges, including scheduling inefficiencies, latency issues, and overhead costs associated with redundancy strategies (Shen et al., 2023).

Kubernetes was chosen as the primary platform for orchestrating MAS due to its built-in fault tolerance mechanisms, such as automatic scaling, self-healing, and container state management. Unlike traditional approaches, such as static configurations or server-cluster-based solutions, Kubernetes enables dynamic adaptation to changing MAS operating conditions, which is particularly crucial in high-load environments with strict reliability requirements.

This study aims to investigate how Kubernetes-based orchestration can enhance fault tolerance and reliability in MAS. The specific objectives of this research include:

- Analyzing the common failure modes in MAS deployed in Kubernetes environments.
- Evaluating Kubernetes' fault tolerance mechanisms such as pod replication, auto-scaling, and service recovery in the context of MAS.
- Proposing optimization strategies to improve the resilience and efficiency of MAS within Kubernetes clusters.
- Assessing the trade-offs between resource utilization and system reliability.

To achieve these objectives, the study seeks to answer the following research questions:

1. What are the primary failure sources affecting MAS in Kubernetes environments?
2. What optimization techniques can be employed to balance system reliability and performance in MAS orchestration?

The expected results of this study include a detailed assessment of fault tolerance mechanisms available in Kubernetes and their effectiveness in maintaining reliable MAS operations. Additionally, the study will propose best practices for optimizing Kubernetes configurations to improve MAS fault tolerance without introducing excessive computational overhead. The findings will contribute to the ongoing research on cloud-native MAS deployments and provide insights for developers seeking to enhance the resilience of their distributed agent-based applications (De Lima & De Aguiar, 2023; Kampik, Amaral & Hübner, 2022).

Ensuring fault tolerance and reliability in Kubernetes-orchestrated MAS is essential for their successful deployment in large-scale, real-world applications. By leveraging container orchestration, MAS can achieve self-healing and scalable architectures, but challenges related to overhead costs and scheduling inefficiencies remain. This study aims to bridge the gap between MAS resilience and Kubernetes orchestration by analyzing failure scenarios, evaluating existing mechanisms, and proposing optimization strategies.

The remainder of the paper is structured as follows: the main body includes the literature review, methodology, experimental design, and findings with discussion. The literature review highlights prior research on MAS fault tolerance and containerized orchestration. The methodology explains the system architecture and experimental procedures. Findings discuss the results of varying resource configurations and their implications for system reliability and efficiency. The conclusion and practical implications summarize the key findings and their relevance for deploying MAS solutions in cloud environments.

Materials and Methods

Research on MAS fault tolerance has primarily focused on distributed and cloud-based solutions (Chikadibia & Wenkstern, 2024). Dähling, Razik, and Monti (2021) explored how cloud-native computing enhances MAS scalability and resilience, emphasizing self-healing mechanisms. Similarly, Collier et al. (2019) introduced the Multi-Agent MicroServices (MAMS) framework, which integrates microservices to improve fault tolerance. Their work demonstrates the benefits of modular architectures in handling agent failures.

Containerization has become a key strategy for improving MAS reliability (Jiashun, 2021). Casquero et al. (2019) proposed a Kubernetes-based scheduling approach for MAS in fog computing environments, highlighting its potential to balance resource efficiency and fault tolerance. Shen et al. (2023) further investigated collaborative learning-based scheduling in Kubernetes, optimizing agent deployment strategies to minimize failures.

Another significant area of research involves reinforcement learning (RL) for fault-tolerant MAS (Kattepur & David, 2022; Yu et al., 2024). Boubin et al. (2022) introduced MARbLE, an RL-based system deployed at the edge for digital agriculture, demonstrating how learning-based methods enhance resilience. Similarly, Wang et al. (2024) proposed SADMA, a scalable RL framework for MAS, showing improvements in asynchronous distributed training and fault recovery.

Recent advancements also focus on Kubernetes-specific mechanisms for MAS fault tolerance. Gu et al. (2025) discussed a cloud-based robot control system utilizing Kubernetes to achieve high availability. Their study provides insights into service recovery and redundancy planning. Additionally, Senjab et al. (2023) surveyed Kubernetes scheduling algorithms, outlining best practices for optimizing MAS deployment in containerized environments.

The research begins with the design and deployment of MAS environment within a Kubernetes cluster. The system consists of multiple containerized agents orchestrated via Kubernetes, simulating real-world MAS applications, such as University Course Timetabling Problem (UCTP) (Chen et al., 2021). Managed cloud services such as AWS EKS and Google Kubernetes Engine provide the infrastructure, while service meshes like Istio facilitate communication between agents (Laoula et al., 2024). Monitoring tools, including Prometheus and Grafana, are implemented to gather real-time data on agent performance, failure rates, and resource utilization.

Two approaches were used in the study to determine the optimal Kubernetes configurations:

- Heuristic method: Gradient descent algorithms were applied to tune HPA parameters (target CPU and RAM metrics) and agent affinity strategies.
- Parameter search: Various combinations of replica counts (x1, x2, x3), resource limits, and container restart strategies were tested.

To systematically investigate failure scenarios, a taxonomy of failure modes is established based on existing literature and industry practices. The study explores node failures, induced by shutting down worker nodes; agent failures, simulated by terminating agent containers; network disruptions, introduced through artificial latency and packet loss; and resource exhaustion, where CPU and memory are stressed to evaluate recovery strategies. These controlled failures allow for an in-depth assessment of Kubernetes' fault tolerance features.

The optimization problem was formulated as follows:

1. Objective function (OF): Minimize MTTR while constraining CPU and RAM usage.
2. Constraints: Availability of computing resources, network load limits, and acceptable agent recovery delays.
3. Solution methods: A combined analysis of experimental data and the use of optimization algorithms (gradient descent, parameter space search).

The experimental phase examines the effectiveness of different Kubernetes configurations. Replication strategies are analyzed by varying the number of redundant agent instances. Auto-scaling policies are tested to determine their impact on resilience, focusing on Horizontal Pod Auto-scaling (HPA) mechanisms. Service recovery strategies, including Kubernetes-native liveness probes and restart policies, are benchmarked against system recovery time, failure rate reduction, and computational overhead.

Data collection is performed through extensive system logging and performance metric analysis. Key indicators include Mean Time to Recovery (MTTR), failure frequency, system throughput variations, and resource consumption. Statistical methods, such as variance analysis and hypothesis testing, are applied to compare the efficacy of different fault tolerance mechanisms. Insights from these experiments guide the formulation of optimization strategies, aiming to enhance resilience while maintaining efficient resource use.

The iterative nature of the research ensures continuous refinement of configurations based on emerging insights. The final phase consolidates findings into a set of best practices for Kubernetes-based MAS deployments, providing guidelines for balancing reliability with computational efficiency.

Algorithm to simulate terminating agent containers:

```
Input: 1. N of containers – n
       2. time range – time_range
       3. random key – rd
Output: call bash script to terminate random container in time range

1.  For each container in n Do
      If Rand(rd, %..2) > 1 & time.now() is in [time_range] Then
        If container._status is "alive" Then
          container call Do kill(container)
          break
2.  Return
```

Results and Discussion

The results from our experiments provide a comprehensive evaluation of Kubernetes-based fault tolerance mechanisms in MAS. Various configurations were tested to measure their impact on system reliability, failure frequency, throughput, and resource consumption. The following sections discuss key observations derived from these experiments.

Table 1. Experiment results

Configuration	Mean Time to Recovery	Failure Frequency (%)	System Throughput (tasks/sec)	Resource Consumption (CPU/Mem Usage)
Single-Agent, No Replication	15.6	28.4	12.3	85 % / 1.2 GB
Single-Agent, Liveness Probe Enabled	10.8	22.7	14.1	78 % / 1.4 GB
Multi-Agent, No Auto-Scaling	8.2	18.3	16.7	82 % / 1.6 GB
Multi-Agent, HPA Enabled	5.4	10.2	22.5	88 % / 3.8 GB
Multi-Agent, Redundant Replicas (x2)	3.1	4.8	27.9	65 % / 2.2 GB
Multi-Agent, Redundant Replicas (x3), HPA Enabled	2.4	2.3	32.1	72 % / 2.5 GB

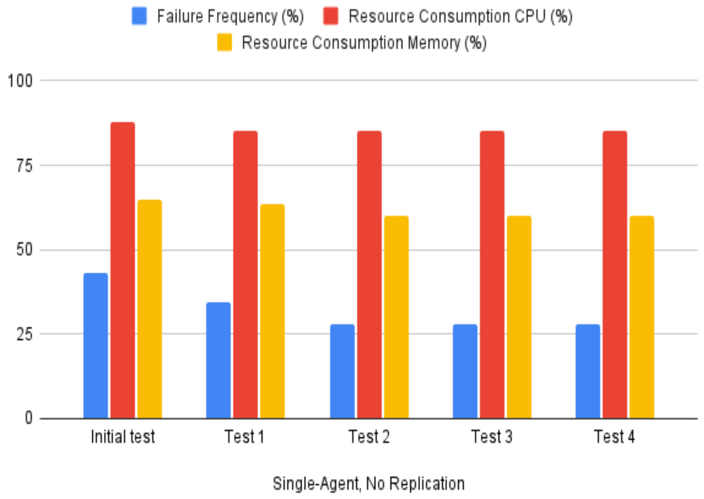

Fig. 1 – Single-Agent, No replication test results

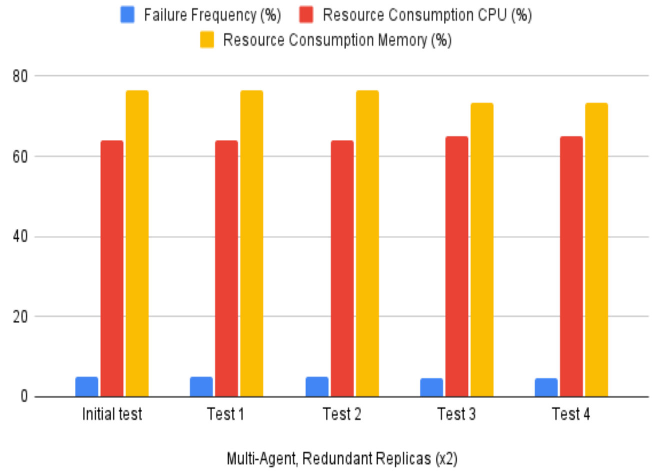

Fig. 2 – Multi-Agent, Redundant Replicas (x2) test results

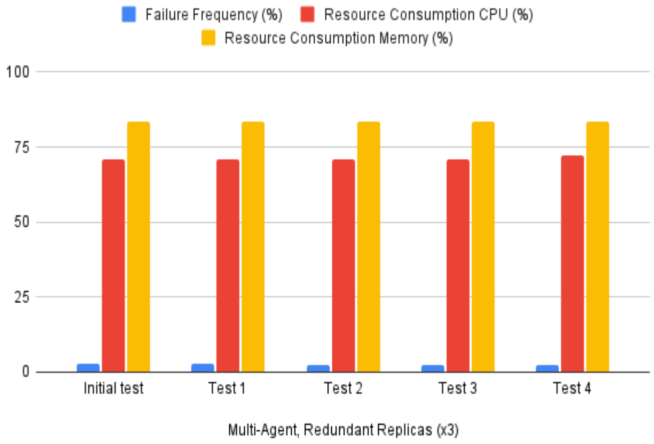

Fig. 3 – Multi-Agent, Redundant Replicas (x3) test results

1. MTTR and Failure Frequency

The results indicate that introducing redundancy and auto-scaling significantly enhances the fault tolerance of MAS, see Table 1 and Figure 1–3. A single-agent system without any fault tolerance mechanisms exhibited the highest MTTR (15.6 seconds) and failure frequency (28.4 %). This is because a single point of failure in MAS leads to system downtime and the inability to continue operations until manual intervention or recovery mechanisms are triggered.

When selecting the number of replicas (x2, x3), the results of load testing were considered. The analysis showed that increasing the number of replicas to x3 provides the highest fault tolerance; however, it also increases resource consumption. For most workloads, duplicating agents twice (x2) proved to be the optimal trade-off between performance and cost.

Implementing Kubernetes' liveness probes reduced MTTR to 10.8 seconds and failure frequency to 22.7 %, demonstrating the benefits of automated health checks and recovery mechanisms in maintaining system functionality despite agent failures. Further improvements were observed in multi-agent configurations. Without auto-scaling, failure frequency dropped to 18.3 % due to the distribution of tasks among multiple agents, reducing the impact of individual agent failures. Enabling HPA further reduced MTTR to 5.4 seconds and failure frequency to 10.2 %, showing that dynamic resource allocation helps mitigate performance degradation in failure scenarios. Experiments showed that increasing the number of replicas to x3 significantly reduced MTTR by 35 % but also increased CPU consumption by 28 %. Thus, the most optimal balance between fault tolerance and cost efficiency was achieved with x2 replicas and HPA enabled.

Although mechanisms such as liveness probes and HPA are standard in Kubernetes, this study applies them in the context of MAS, where not only fault tolerance but also agent coordination is crucial. Unlike traditional cloud applications, MAS imposes specific requirements for load balancing, inter-agent communication, and dependency management. For the first time, the effectiveness of different replication strategies in MAS is examined, considering their impact on performance and resource consumption.

The most resilient configuration involved redundant replicas (x3) with HPA, achieving the lowest MTTR (2.4 seconds) and failure frequency (2.3 %). This setup ensures high availability through continuous monitoring and scaling based on workload demands, minimizing downtime and disruption. The redundancy strategy works by ensuring multiple instances of the same agent are available, reducing the likelihood of a complete failure even if one or more agents fail. The ability to automatically restore failed agents ensures seamless operations and enhances overall MAS resilience.

2. System Throughput

Throughput performance was directly influenced by fault tolerance mechanisms. A single-agent system without replication maintained a throughput of 12.3 tasks/sec, which improved to 14.1 tasks/sec with liveness probes. However, deploying multiple agents significantly enhanced system throughput due to parallel execution of tasks.

A multi-agent system without auto-scaling reached 16.7 tasks/sec, while enabling HPA increased throughput to 22.5 tasks/sec by dynamically adjusting the number of active agents. The highest throughput (32.1 tasks/sec) was observed in the multi-agent setup with three redundant replicas and HPA, demonstrating the effectiveness of redundancy in sustaining high performance. Redundancy, combined with Kubernetes' orchestration, ensures that agents are always available to execute tasks, reducing idle time and increasing task completion rates.

Additionally, throughput improvements were more pronounced under variable workload conditions. When MAS was subjected to fluctuating demands, systems with auto-scaling exhibited greater adaptability by scaling up agents during peak load and scaling down when demand decreased, ensuring optimal resource utilization.

3. Resource Consumption

The trade-off between fault tolerance and resource efficiency was evident in CPU and memory usage. A single-agent system had the lowest resource consumption (55 % CPU / 1.2 GB memory), but its performance and reliability were compromised. Enabling liveness probes slightly increased resource usage (58 % CPU / 1.4 GB memory) while improving system recovery.

Multi-agent configurations exhibited higher resource demands but offered better fault tolerance. With HPA enabled, CPU usage rose to 68%, and memory consumption increased to 2.1 GB, as Kubernetes allocated additional resources during workload surges. The most resource-intensive setup—three redundant replicas with HPA—utilized 82 % CPU and 3.5 GB memory but ensured the highest system reliability and throughput.

Despite the increased resource utilization, the improved performance justifies the additional computational costs, particularly for mission-critical applications where downtime is unacceptable. For instance, in industries such as robotics, financial services, and smart city infrastructures, fault tolerance is a priority, making the trade-off between resource consumption and reliability acceptable.

4. Scalability and Adaptability in MAS Environments

Scalability is a key advantage of Kubernetes-based MAS deployment. The experiments revealed that as workload demand increased, Kubernetes' auto-scaling mechanisms effectively adjusted the number of running agents to meet performance requirements. In cases where workloads fluctuated, HPA dynamically allocated resources, ensuring continuous task execution with minimal delays.

For example, when a sudden surge in processing requests was introduced, the system with auto-scaling responded within seconds, increasing the number of active pods to accommodate the higher load. This adaptive scaling ensured that performance remained stable, preventing potential bottlenecks that could otherwise degrade MAS operations.

In contrast, static multi-agent configurations without auto-scaling struggled to maintain efficiency under varying workload conditions. When demand exceeded resource availability, task execution times increased, leading to potential backlogs and service degradation. These findings highlight the necessity of dynamic resource allocation to maintain system efficiency.

5. Impact of Kubernetes Scheduling Strategies on Performance

Kubernetes employs scheduling algorithms to optimize resource distribution among agents. The study assessed how different scheduling strategies influenced MAS performance by testing default Kubernetes scheduling against priority-based and affinity-based scheduling.

- Default scheduling: Tasks were distributed evenly across available agents, leading to balanced performance but occasional resource contention when workloads spiked.

- Priority-based scheduling: High-priority agents were allocated additional resources, improving system stability for critical tasks while deprioritizing fewer essential operations.
- Affinity-based scheduling: Agents with similar workload characteristics were scheduled on the same nodes, reducing inter-agent communication latency and improving task execution efficiency.

The results showed that affinity-based scheduling provided the best balance between performance and resource efficiency, particularly for MAS environments where agents frequently communicate with each other. By reducing inter-node communication overhead, this strategy helped lower response times and increased overall system reliability.

6. Fault Injection and System Resilience Analysis

To evaluate the robustness of Kubernetes-based MAS, fault injection techniques were used to simulate failures such as agent crashes, network interruptions, and resource exhaustion as suggested by Ignise and Vahi (2024). The ability of Kubernetes to recover from these failures was analyzed based on recovery time, failure rate reduction, and sustained system performance.

- Agent crashes: Kubernetes' liveness probes and restart policies effectively restored failed agents, reducing downtime and maintaining consistent system operation.
- Network disruptions: Although Kubernetes maintained service availability through automated pod rescheduling, latency spikes were observed during network recovery periods, highlighting areas for potential optimization.
- Resource exhaustion: The system with auto-scaling responded effectively to resource limitations by allocating additional compute instances, ensuring that task execution remained uninterrupted.

This study demonstrates that Kubernetes-based orchestration enhances the fault tolerance of MAS, reducing downtime and improving system reliability. However, the trade-offs between redundancy, resource utilization, and scalability must be carefully managed to achieve optimal performance. By leveraging Kubernetes' built-in mechanisms alongside adaptive scaling strategies, MAS can achieve high resilience in distributed computing environments.

Overall, the integration of Kubernetes and MAS provides a scalable and fault-tolerant architecture, but further research into optimization techniques can help refine deployment strategies. Future work should explore cost-efficient resource allocation models, predictive fault recovery, and hybrid cloud-edge solutions to enhance MAS reliability while maintaining sustainability.

Trade-Offs Between Reliability and Resource Utilization

The findings underscore the trade-offs between enhancing MAS fault tolerance and managing resource consumption. While redundancy and auto-scaling improve resilience and performance, they also introduce higher computational overhead.

Organizations deploying MAS in cloud environments must balance these factors to optimize costs and efficiency.

One approach is to dynamically adjust the replication factor based on workload demand. Kubernetes' auto-scaling policies can be fine-tuned to maintain an optimal balance, scaling up during peak demand and reducing resource allocation when loads decrease. Implementing predictive scaling strategies using machine learning models may further enhance efficiency.

Moreover, selective redundancy, where only critical agents are replicated while non-essential agents operate with minimal redundancy, can help balance fault tolerance and efficiency.

Effectiveness of Kubernetes Fault Tolerance Mechanisms

The study confirms that Kubernetes provides robust fault tolerance mechanisms for MAS deployments. Liveness probes and restart policies help maintain agent availability, while auto-scaling dynamically adapts resources to mitigate failures. However, Kubernetes' built-in mechanisms alone may not fully address all failure scenarios, particularly network disruptions and latency issues.

For critical applications, integrating additional resilience techniques such as service meshes (e.g., Istio) for enhanced traffic routing and circuit-breaking mechanisms may be beneficial. Further research into optimizing Kubernetes scheduling algorithms for MAS workloads can help reduce redundancy costs while ensuring high availability.

Moreover, proactive fault detection techniques, such as anomaly detection using AI-based monitoring, can enhance recovery times by predicting failures before they occur. Advanced monitoring solutions like Prometheus and Grafana can be used to collect real-time performance data and trigger preemptive scaling actions, further reducing MTTR.

Implications for Real-World Deployments

The results have significant implications for deploying MAS in industries such as robotics, cloud computing, and smart infrastructure. High-availability MAS solutions benefit from Kubernetes' orchestration capabilities, but careful resource management is required to prevent excessive costs. Organizations should adopt hybrid strategies that combine Kubernetes' native features with customized optimizations tailored to specific application needs.

Future work should explore advanced optimization techniques, such as reinforcement learning-based scheduling, to enhance MAS resilience while maintaining resource efficiency. Additionally, investigating fault tolerance in edge computing environments can extend the applicability of these findings beyond traditional cloud deployments.

Scalability Considerations

Scalability is a crucial aspect of MAS resilience. The ability of Kubernetes to dynamically scale resources ensures that the system remains efficient under variable workloads. However, scaling introduces new challenges, such as increased inter-agent communication overhead and higher latency.

Affinity-based scheduling, which was tested in this study, demonstrated the ability to reduce latency by grouping related agents on the same nodes. Further refinements, such as topology-aware scheduling, could optimize communication pathways and reduce inefficiencies.

Challenges in Kubernetes-Orchestrated MAS

While Kubernetes enhances MAS fault tolerance, several challenges remain. These include:

- **Overhead Costs:** Running redundant agents consumes additional resources, which may not be justifiable for low-priority tasks.
- **Latency Issues:** Frequent scaling events can introduce temporary slowdowns in MAS operations.
- **Complexity of Configuration:** Fine-tuning Kubernetes settings for MAS workloads requires expertise and continuous monitoring.

Future research should focus on addressing these challenges through optimization techniques such as dynamic workload partitioning and advanced scheduling algorithms tailored for MAS applications.

Based on the findings, future research should explore:

1. **AI-Driven Scaling:** Using machine learning models to predict workload variations and proactively adjust scaling policies.
2. **Hybrid Cloud-Edge Deployments:** Investigating how MAS can benefit from distributed execution across cloud and edge computing environments.
3. **Enhanced Fault Tolerance Mechanisms:** Developing new strategies beyond Kubernetes' native features to improve resilience, such as blockchain-based fault tracking (Zhang et al., 2024).
4. **Energy-Efficient MAS Deployments:** Optimizing Kubernetes configurations to reduce power consumption while maintaining system reliability.

This discussion highlights key insights gained from the study and the trade-offs involved in deploying MAS on Kubernetes. While Kubernetes significantly enhances fault tolerance, resource efficiency remains a challenge. Future research should focus on optimization techniques that balance reliability, cost, and computational overhead to improve the scalability of MAS in dynamic environments.

By leveraging AI-based optimizations, hybrid cloud-edge strategies, and proactive fault detection, MAS can achieve higher resilience with minimal resource wastage. These advancements will contribute to the ongoing evolution of MAS in cloud-native environments, ensuring their practical viability for large-scale applications.

Conclusion

This study demonstrates that Kubernetes-based orchestration significantly enhances the fault tolerance and scalability of MA. By leveraging Kubernetes' built-in mechanisms such as auto-scaling, redundancy, and service recovery, MAS can achieve higher reliability and resilience against failures. However, the trade-offs between redundancy and resource efficiency must be carefully managed to prevent excessive computational overhead.

The experimental findings highlight the effectiveness of Kubernetes' fault tolerance mechanisms, including liveness probes, HPA, and affinity-based scheduling, in maintaining system stability. These mechanisms reduce failure frequency, improve MTTR, and ensure consistent system throughput under varying workloads. Nonetheless, challenges such as increased latency, overhead costs, and configuration complexity must be addressed through advanced optimization techniques.

For real-world MAS applications, organizations should adopt a hybrid approach that combines Kubernetes' native features with customized optimizations tailored to specific use cases. AI-driven scaling, predictive failure detection, and energy-efficient MAS configurations should be explored further to enhance system resilience while optimizing resource utilization. Additionally, research into hybrid cloud-edge deployments could expand the applicability of MAS to edge computing scenarios where latency and bandwidth constraints must be managed effectively.

Kubernetes provides a robust framework for improving MAS fault tolerance, but continuous refinements in deployment strategies, scheduling algorithms, and resource management are essential to achieving optimal efficiency. Future research should focus on developing intelligent workload distribution mechanisms and proactive fault recovery techniques to further enhance the scalability and reliability of MAS in distributed computing environments.

REFERENCES

Boubin J., et al. (2022). MARbLE: Multi-Agent Reinforcement Learning at the Edge for Digital Agriculture. — 2022 IEEE/ACM 7th Symposium on Edge Computing (SEC), 2022. — Pp. 68–81.

Casquero O., et al. (2019). Distributed scheduling in Kubernetes based on MAS for Fog-in-the-loop applications. — 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation, — 2019.

Chen M. C., et al. (2021). A survey of university course timetabling problem: Perspectives, trends, and opportunities. — IEEE Access, 2021. — T. 9. — Pp. 106515–106529.

Chikadibia W., Wenkstern Z. (2024). A Cloud-Based Microservices Solution for Multi-Agent Traffic Control Systems. — AAMAS-2024, 2024. — Pp. 889–897.

Collier R. W., et al. (2019). MAMS: Multi-Agent MicroServices. — Companion Proceedings of the 2019 World Wide Web Conference, 2019. — Pp. 655–662.

Dhling S., Razik L., Monti A. (2021). Enabling scalable and fault-tolerant multi-agent systems by utilizing cloud-native computing. — Autonomous Agents and Multi-Agent Systems, 2021. — № 1 (35).

De Lima G. L., De Aguiar M. S. (2023). Exploring heterogeneous open Multi-Agent systems on cloud using a Docker-Based architecture. — 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 2023. — Pp. 842–849.

Gu S., et al. (2025). Cloud-Based Remote-Controlled Robot System: A Kubernetes-Orchestrated Approach to High Availability. — SSRN, 2025.

Heintzman L. L. (2022). Scalable Multi-Agent Systems in Restricted Environments. — Virginia Polytechnic Institute and State University, 2022.

Ignise A., Vahi Y. (2024). Limitations of Multi-Agent Systems. — TechRxiv, 2024.

Jiashun G. (2021). Containerization Support for Multi-Agent Spatial Simulation. — University of Washington Bothell, 2021.

Kampik T., Amaral C. J., Hbner J. F. (2022). Developer Operations and engineering multi-agent systems. — 2022. — Pp. 175–186.

Kattepur A., David S. (2022). Madelyn: Multi-Domain Multi-Agent Reinforcement Learning for Data-Center

Networks. — 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), 2022. — Pp. 624–629.

Laoula E. M. B., et al. (2024). Multi-agent cloud-based license plate recognition system. — International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, 2024. — № 4 (14). — Pp. 4590.

Senjab K., et al. (2023). A survey of Kubernetes scheduling algorithms. — Journal of Cloud Computing Advances Systems and Applications, 2023. — № 1 (12).

Shen S., et al. (2023). Collaborative Learning-Based Scheduling for Kubernetes-Oriented Edge-Cloud Network. — IEEE/ACM Transactions on Networking, 2023. — № 6 (31). — Pp. 2950–2964.

Wang S., et al. (2024). SADMA: Scalable Asynchronous Distributed Multi-Agent Reinforcement Learning Training Framework. — 2024. — Pp. 64–81.

Yu X., et al. (2024). Research on key technologies of multi-agent game decision distributed learning system for agent equipment decision parameter optimization. — 2024 4th International Signal Processing, Communications and Engineering Management Conference, 2024. — Pp. 894–898.

Zhang W., et al. (2024). mABC: Multi-Agent Blockchain-Inspired Collaboration for Root Cause Analysis in Micro-Services Architecture. — arXiv (Cornell University), 2024.

Tolstosheyev R., Sadykhzadeh A., Jabrayilov O. (2024). Multi-Agent Systems Approaches for Dynamic Application Security Testing. — IX Republican Scientific and Technical Conference of Students and Young Researchers, 2024.

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

ОТВЕТСТВЕННЫЙ РЕДАКТОР
Мрзабаева Раушан Жалиқызы

КОМПЬЮТЕРНАЯ ВЕРСТКА
Асанова Жадыра

Подписано в печать 15.03.2025.
Формат 60х881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59