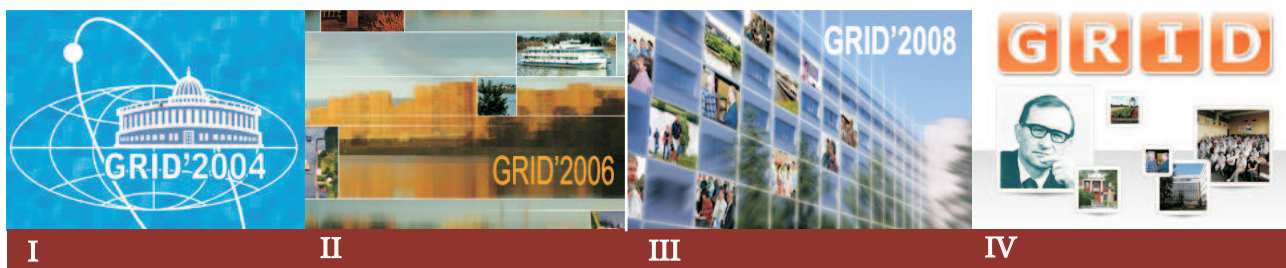




**DISTRIBUTED COMPUTING
AND GRID-TECHNOLOGIES IN SCIENCE
AND EDUCATION**

*Proceedings of the 5th International Conference
Dubna, 16 - 21 July, 2012*



**РАСПРЕДЕЛЕННЫЕ ВЫЧИСЛЕНИЯ
И ГРИД-ТЕХНОЛОГИИ В НАУКЕ
И ОБРАЗОВАНИИ**

*Труды 5-й международной конференции
Дубна, 16 - 21 июля, 2012 г.*

GRID'2012



**JOINT INSTITUTE FOR NUCLEAR RESEARCH
LABORATORY OF INFORMATION TECHNOLOGIES**

**Distributed Computing and
Grid-technologies in
Science and Education**

Proceedings of 5th International Conference

Dubna, 16 – 21 July, 2012

**Распределенные вычисления и
Грид-технологии в
науке и образовании**

Труды 5-ой международной конференции

Дубна, 16 – 21 июля 2012 г.

ДУБНА, 2012

УДК[004.7+004.9](063)
ББК[32.988-5я431+32.97я431]
D63

Organized by the Joint Institute for Nuclear Research,
The Laboratory of Information Technologies
under the Sponsorship of the Russian Foundation for Basic Research,
Supermicro Computer,
NIAGARA
Media partners
PARALLEL.RU

Distributed Computing and Grid-Technologies in Science and Education: Proceedings of the 5th Intern. Conf. (Dubna, 16-21 July, 2012). – Dubna: JINR, Д-11-2012-127, 2012.-p.419
ISBN 978-5-9530-0345-2

The Proceedings of the 5th International Conference “Distributed Computing and Grid-Technologies in Science and Education” (GRID’2012) include the reports presented at the GRID’2012, which was held in Dubna on 16 – 21 July, 2012. The Conferences are held every two years by the JINR Laboratory of Information Technologies. The Conference is a unique one conducted in Russia on the issues relating to the use of Grid-technologies in various areas of science, education, industry and business. The tentative timetable of this conference foresees in-depth analysis of hot topics such as the development of the global grid-infrastructure providing the optimal use of computing and data storage resources through national grid-infrastructures (National Grid Initiatives, NGIs) as well as the development of the Russian Grid Network. Special sections have been devoted to the emerging direction of the cloud computing and the desktop grids. Presentations of the Conference are available at the Conference web-page <http://grid2012.jinr.ru/program.php>

Распределенные вычисления и Грид-технологии в науке и образовании: Труды 5-й междунар. конф. (Дубна, 16 -21 июля 2012 г.)- Дубна: ОИЯИ, Д-11-2012-127, 2012.-с.419
ISBN 978-5-9530-0345-2

Труды 5-ой международной конференции «Распределенные вычисления и Грид-технологии в науке и образовании» (GRID’2012) содержат доклады, представленные на GRID’2012, которая проходила 16 -21 июля 2012 г. в г.Дубне. Конференция проводится раз в два года Лабораторией информационных технологий ОИЯИ. Конференция является единственной в России, посвященной проблемам, связанным с использованием грид-технологий в различных областях науки, образования, промышленности и бизнеса. На конференции было проанализировано развитие глобальной грид-инфраструктуры, обеспечивающей оптимальное использование вычислительных ресурсов и ресурсов хранения данных, на основе национальных грид-инфраструктур (National Grid Initiatives, NGIs) и развитие Российской грид-сети. Специальные секции были посвящены «облачным» вычислениям и грид-системам из персональных компьютеров. Презентации докладов, представленные на конференции размещены на веб-странице конференции <http://grid2012.jinr.ru/program.php>

УДК[004.7+004.9](063)
ББК[32.988-5я431+32.97я431]

ISBN -5-9530-0345-2

©Joint Institute for Nuclear research, 2012

General Information

The 5th International Conference "Distributed Computing and Grid-technologies in Science and Education" was held from 16 – 21 July, 2012 in the Laboratory of Information Technologies of the Joint Institute for Nuclear Research (Dubna, Russia). This event was the fifth one held by LIT in this subject area every two years. The program of the Conference included not only the questions related to creation and operation of Grid-infrastructures and Grid-applications, but also some theoretical and practical aspects of utilizing distributed computing environments, distributed data processing, etc. This time the heightened interest to the Conference was aroused by the creation in Russia of a Tier1 level data-processing centre at JINR and at NRC "Kurchatov Institute" as well as by vigorous activity in applying the so-called "cloud computing". The Conference was attended by 256 participants from 22 countries: Azerbaijan, Belarus, Bulgaria, Great Britain, Germany, Georgia, Italy, Kazakhstan, China, Cuba, Moldova, Mongolia, Myanmar, Russia, Romania, USA, Uzbekistan, Ukraine, France, Czechia, Switzerland, Sweden as well as CERN and JINR. Russia was presented by participants from 40 universities and research centers. The Conference program included daily plenary sessions and 8 sections: Grid-infrastructures, Clouds and Grid, Grid-applications, Desktop grids, Systems of distributed information resources, WLCG (Worldwide LHC Computing Grid), GridNNN (Grid of the National Nanotechnology Network), Distributed computing: methods and algorithms as well as poster presentations. A workshop on computing for ATLAS experiment and a round-table discussion devoted to the creation of a Tier-1 level data processing centre for LHC experiments in Russia were organized in frames of the Conference.

Advisory Committee

Abdinov O. (IoP, Baku, Republic of Azerbaijan), Afanasiev A.P. (ISA RAS, Moscow, Russian Federation), Antoniou I. (Aristotle University of Thessaloniki, Greece), Andreeva J. (CERN), Bird I. (CERN), Bogatencov P. (RENAM, Moldova), Bogdanov A.V. (IHPCIS, St.Petersburg, Russian Federation), Burtebaev N. (Institute of Nuclear Physics, Kazakhstan), Buša J. (TU, Košice, Slovakia), Buzatu F. (Institute for Atomic Physics, Magurele, Romania), Chetverushkin B.N. (Keldysh Institute of Applied Mathematics, Moscow), De K. (University of Texas at Arlington, USA), Dimitrov V. (Sofia University, Republic of Bulgaria), Dulea M. (IFIN-HH, Romania), Elizbarashvili A. (Tbilisi State University, Georgia), Gusev V.V. (IHEP, Protvino, Russian Federation), Ilyin V.A. (SINP MSU / NRC "Kurchatov Institute", Moscow, Russian Federation), Kitowski J. (CYFRONET, Republic of Poland), Klementov A. (BNL, USA), Kryukov A.P. (SINP MSU / NRC "Kurchatov Institute", Moscow, Russian Federation), Kudrov N.I. (Ministry of Communications and Mass Media of the Russian Federation), Kudryashov N.A. (NRNU MEPhI, Moscow, Russian Federation), Lakhno V.D. (IMPB Russian Academy of Sciences, Russian Federation), Lokajicek M. (PI AS CzR, Czech Republic), Matveev V.A. (JINR), Metakides G. (University of Patras, Greece), Manh Shat Nguyen (JINR), Musial G. (Institute of Physics, AMU, Poznan, Poland), Nergui B. (Institute of Informatics MAS, Mongolia), Ratnikova N. (Karlsruhe Institute of Technology, Germany), Ryabov Yu.F. (PNPI, Gatchina, Russian Federation), Sahakyan V.G. (IIAP NAS, Armenia), Savin G.I. (JSCC RAS, Moscow, Russian Federation), Shumeiko N.M. (NC PHEP, Minsk, Republic of Belarus), Smirnova O.G. (NDGF/University of Lund, Sweden), Vaniachine A.V. (Argonne National Laboratory, USA), Velikhov V.E. (NRC "Kurchatov Institute", Moscow, Russian Federation), Voevodin V.V. (SRCC Moscow State University, Russian Federation), Zhizhin M.N. (CGDS RAS, Russian Federation), Zhizhchenko A.B. (JSCC RAS, Moscow, Russian Federation), Zinovjev G. (ITP, Kiev, Ukraine)

Organizing Committee (JINR)

Ivanov V.V. - Chairman, Korenkov V.V. - Vice-Chairman, Strizh T.A. - Scientific Secretary, Adam S., Aristarkhova M.V., Bulyga N.I., Fedorova E.A., Grafov A.N., Katraseva T.I., Korotchik O.M., Lukyanov S.O., Podgainy D.V., Prikhodko A.V., Rudneva E.M., Rumyantseva O.Yu., Streltsova O.I., Tikhonenko E.A., Torosyan Sh., Zrellov P.V.

Общая информация

Международная конференция "Распределенные вычисления и Грид-технологии в науке и образовании", проводимая раз в два года Лабораторией информационных технологий, проходила в Объединенном институте ядерных исследований с 16 по 21 июля 2012 г. Настоящая конференция была пятой по данной тематике, проводимой ЛИТ раз в два года. Программа конференции включала не только вопросы, связанные с созданием и эксплуатацией грид-инфраструктур и грид-приложений, но и теоретические и практические аспекты применения распределенных вычислительных сред, распределенной обработки данных и др. В этом году повышенный интерес к конференции был связан с созданием в России центра обработки данных уровня Tier1 на базе ОИЯИ и НИЦ «Курчатовский институт» и активной деятельностью в области применения «облачных вычислений». Конференция собрала 256 участника из 22 стран: Азербайджана, Белоруссии, Болгарии, Великобритании, Германии, Грузии, Италии, Казахстана, Китая, Кубы, Молдавии, Монголии, Мьянмы, России, Румынии, США, Узбекистана, Украины, Франции, Чехии, Швейцарии, Швеции, а также ЦЕРН и ОИЯИ. Россия была представлена участниками из 40 университетов и исследовательских центров. На конференции была организована работа ежедневных пленарных заседаний и 8 секций: Грид-инфраструктуры, «облачные» вычисления и грид, Грид-приложения, Грид-системы из персональных компьютеров, системы распределенных информационных ресурсов, WLCG – Всемирный грид для обработки данных с Большого адронного коллайдера в ЦЕРН, ГридННС – грид национальной нанотехнологической сети, распределенные вычисления (методы и алгоритмы) и стендовые доклады. В рамках конференции было проведено рабочее совещание по компьютерингу для эксперимента ATLAS и круглый стол, посвященный созданию в России центра обработки данных уровня Tier-1 для экспериментов на БАК.

Программный комитет

Абдинов О. (Институт физики, Баку, Азербайджанская Республика), Афанасьев А.П. (ИСА РАН, Москва, Россия), Антониу И. (Университет имени Аристотеля, Салоники, Греция), Андреева Ю. (ЦЕРН), Берд Я. (ЦЕРН), Богатенков П. (RENARM, Молдова), Богданов А.В. (ИВВИС, Санкт-Петербург, Россия), Буртебаев Н. (Институт ядерной физики, Казахстан), Буша Я. (Технический университет, Кошице, Словакия), Бузату Ф. (IFIN-НН, Румыния), Ваняшин А.В. (Аргоннская национальная лаборатория, США), Велихов В.Е. (НИЦ «Курчатовский институт, Россия), Воеводин В.В. (НИВЦ МГУ, Россия), Гусев В.В. (ИФВЭ, Протвино, Россия), Де К. (Техасский университет, Арлингтон, США), Димитров В. (Софийский университет, Республика Болгария), Дулеа М. (IFIN-НН, Румыния), Жижин М.Н. (Центр геофизических данных РАН, Россия), Жижченко А.Б. (Межведомственный Суперкомпьютерный Центр, Россия), Зиновьев Г. (ИТФ, Киев, Украина), Ильин В.А. (НИИЯФ МГУ/НИЦ «Курчатовский институт, Россия), Китовски Я. (CIFRONET, Республика Польша), Климентов А. (БНЛ, США), Крюков А.П. (НИИЯФ МГУ/НИЦ «Курчатовский институт, Россия), Кудров Н.И. (Минкомсвязь, Россия), Кудряшов Н.А. (НИЯУ «МИФИ», Москва, Россия), Лахно В.Д. (ИМПБ РАН, Россия), Локайчик М. (Институт физики, Чехия), Матвеев В.А. (ОИЯИ), Метакидис Г. (Университет Патрас, Греция), Мусял Г. (ИФ Унив. Им. А.Мицкевича, Познань, Польша), Нгуен Мань Шат (ОИЯИ), Нэргуй Б. (Институт информатики МАН, Монголия), Ратникова Н. (Институт технологий Карлсруэ, Германия), Рябов Ю.Ф. (ПИЯФ РАН, Россия), Саакян В.Г. (ИПИА, Армения), Савин Г.И. (Межведомственный Суперкомпьютерный Центр, Россия), Смирнова О.Г. (NDGF/Университет Лунда, Швеция), Четверушкин Б.Н. (ИПМ им. М.В.Келдыша, Москва, Россия), Шумейко Н.М. (ИЦ ФЧВЭ БГУ, Минск, Белоруссия), Элизбарашвили А. (Тбилисский государственный университет, Грузия).

Организационный комитет (ОИЯИ)

Иванов В.В. - председатель, Кореньков В.В. - зам. председателя, Стриж Т.А. - ученый секретарь, Адам С., Аристархова М.В., Булыга Н.И., Графов А.Н., Зрелов П.В., Катрасева Т.И., Коротчик О.М., Лукьянов С.О., Подгайный Д.В., Приходько А.В., Руднева Е.М., Румянцева О.Ю., Стрельцова О.И., Тихоненко Е.А., Торосян Ш., Федорова Е.А.



CONTENTS / СОДЕРЖАНИЕ

IMAGE PROCESSING BY CORE CLUSTERIZATION ALGORITHM.....	13
F.T. Adilova, R.R. Davronov	
GRID TECHNOLOGIES IN SPbSU LONG-RANGE CORRELATIONS ANALYSIS AND MC SIMULATIONS FOR ALICE	18
I.G. Altsybeev, G.A. Feofilov, M.V. Kompaniets, V.N. Kovalenko, V.V. Vechernin, I.S. Vorobyev, A.K. Zarochentsev	
GRID IN JINR AND PARTICIPATION IN THE WLCG PROJECT	23
S.D. Belov, P.V. Dmitrienko, V.V. Galaktionov, N.I. Gromova, I.S. Kadochnikov, V.V. Korenkov, N.A. Kutovskiy, S.V. Mitsyn, V.V. Mitsyn, D.A. Oleynik, A.S. Petrosyan, G.S. Shabratova, T.A. Strizh, E.A. Tikhonenko, V.V. Trofimov, V.E. Zhiltsov, A.V. Uzhinskiy	
MONITORING, ACCOUNTING AND REGISTRATION SERVICES FOR RUSSIAN GRID NETWORK	30
S.D. Belov, T.M. Goloskokova, V.V. Korenkov, N.A. Kutovskiy, D.A. Oleynik, A.S. Petrosyan, R.N. Semenov, A.V. Uzhinskiy	
SUPPORT FOR THE CMS EXPERIMENT AT THE TIER-1 CENTER IN GERMANY	34
J. Berger, C. Böser, T. Chwalek, M. Fischer, O. Oberst, G. Quast, N. Ratnikova, S. Röcker, S. Wayand, M. Zeise, M. Zvada	
A DISTRIBUTED BRANCH AND BOUND METHOD FOR BOINC DESKTOP GRIDS	43
Bo Tian, M. Posypkin	
PRACTICAL EFFICIENCY OF OPTIMIZING COMPILERS IN PARALLEL SCIENTIFIC APPLICATIONS	48
A.V. Bogdanov, I.G. Gankevich	
VIRTUALIZATION WITH ORACLE SOLARIS 10	54
A.V. Bogdanov, Pyae Sone Ko Ko	
PRIVATE CLOUD VS PERSONAL SUPERCOMPUTER	57
A.V. Bogdanov	
VIRTUAL WORKSPACE AS BASIS OF SUPERCOMPUTER CENTER	60
A.V. Bogdanov, A.B. Degtyarev, V.Yu. Gaiduchok, I.G. Gankevich, V.I. Zolotarev	
CPU AND GPU CONSOLIDATION BASED ON OPENCL	66
A.V. Bogdanov, I.G. Gankevich, V.Yu. Gayduchok, Pyae Sone Ko Ko	
SCALING THE SPEEDUP OF MULTI-CORE CHIPS BASED ON AMDAHL'S LAW	71
A.V. Bogdanov, Kyaw Zaya	
DATABASE CONSOLIDATION USED FOR PRIVATE CLOUD	76
A.V. Bogdanov, Thurein Kyaw Lwin, Ye Myint Naing	
IMPROVING THE EFFICIENCY OF DISTRIBUTED INTELLIGENT SYSTEMS	81
A.B. Degtyarev, V.P. Guskov, A.V. Eroshkin	

BES-III DISTRIBUTED COMPUTING	85
Z.Y. Deng, W.D. Li, L. Lin, C. Nicholson, X.M. Zhang, A. Zhemchugov	
CERTREQ: A STANDALONE TOOL FOR CERTIFICATE REQUESTS GENERATION AND CERTIFICATES RETRIEVING IN GRIDNNN	89
Yu.Yu. Dubenskaya, A.P. Kryukov, L.V. Shamardin	
GRID AND HPC SUPPORT FOR NATIONAL PARTICIPATION IN LARGE-SCALE COLLABORATIONS	95
M. Dulea, Ș. Constantinescu, M. Ciubăncan, T. Ivănoaica, C. Plăcintă, I.T. Vasile, D. Ciobanu-Zabet	
CONFIGURATION MANAGEMENT FOR IT INFRASTRUCTURE	104
O. Dulov	
ATLAS TIER 3 IN GEORGIA	111
A. Elizbarashvili	
TECHNOLOGY OF SEMANTIC STRUCTURING OF THE DIGITAL LIBRARY CONTENT	117
I.A. Filozova	
APPLICATION OF DESKTOP GRID TECHNOLOGY IN MATERIAL SCIENCE	123
O. Gatsenko, O. Baskova, B. Bandalak, V. Tatarenko, Yu. Gordienko	
DEVELOPMENT OF THE VO-SPECIFIC dCache DATA BROWSER	130
M. Gavrilenko, I. Gorbunov, V. Korenkov, D. Oleynik, A. Petrosyan, S. Shmatov	
CMS EXPERIMENT DATA PROCESSING AT RDMS CMS TIER 2 CENTERS	133
V. Gavrilov, I. Golutvin, V. Korenkov, E. Tikhonenko, S. Shmatov, V. Zhiltsov, V. Ilyin, O. Kodolova, L. Levchuk	
ARCHITECTURE OF A SOA-BASED BPM PLATFORM FOR THE EGI	138
R.D. Goranova	
MODEL OF DATA STORAGE AND PROCESSING SYSTEM FOR “PIK” NUCLEAR REACTOR EXPERIMENTS	144
A.P. Gulin, A.K. Kiryanov, N.V. Klopov, E.G. Novodvorsky, S.B. Oleshko, Y.F. Ryabov	
ON APPROACHES TO BUILDING PROBLEM-ORIENTED WEB-INTERFACES FOR APPLICATION SOFTWARE SUITES IN GRIDNNN	147
A.P. Gulin, A.K. Kiryanov, N.V. Klopov, S.B. Oleshko, Y.F. Ryabov	
WLCG TIER-2 COMPUTING INFRASTRUCTURE AT IHEP	150
V. Gusev, V. Kotlyar, V. Kukhtenkov, E. Popova, N. Savin, A. Soldatov	
OVERALL EXPERIENCE OF GRIDKA T1 OPERATIONS AND LHC EXPERIMENTS REPRESENTATION.....	158
A. Heiss, A. Petzold, M. Zvada	
A STATUS REPORT ON CLUSTERING AND SERVICES DEPLOYED AT ISS DATE CENTER	165
F.L. Irimia, A. Sevcenco, B.A. Dumitru, I. Stan, S. Zgura	

TORRENT BASE OF SOFTWARE DISTRIBUTION BY ALICE AT RDIG	171
V. Kotlyar, E. Ryabinkin, G. Shabratova, I. Tkachenko, A. Zarochentsev	
VIRTUAL ACCELERATOR: GRID-ORIENTED SOFTWARE FOR BEAM ACCELERATOR CONTROL SYSTEM	176
N.V. Kulabukhova, A.N. Ivanov, V.V. Korkhov, D.A. Vasyunin, S.N. Andrianov	
DISTRIBUTED TRAINING AND TESTING GRID INFRASTRUCTURE EVOLUTION	180
N.A. Kutovskiy	
PROBLEM-ORIENTED WEB-INTERFACES FOR RUSSIAN GRID NETWORK	186
N.A. Kutovskiy, I.I. Lensky, R.N. Semenov	
DDM DQ DELETION SERVICE. IMPLEMENTATION OF CENTRAL DELETION SERVICE FOR ATLAS EXPERIMENT	189
D. Oleynik, A. Petrosyan, V. Garonne, S. Campana on behalf of the ATLAS Collaboration	
ATLAS OFF-GRID SITES (TIER-3) MONITORING	195
A. Petrosyan, D. Oleynik, S. Belov, J. Andreeva, I. Kadochnikov on behalf of the ATLAS Collaboration	
META-MONITORING WITH THE HAPPYFACE PROJECT	200
S. Röcker, A. Burgmeier, M. Heinrich, G. Quast, G. Vollmer, M. Zvada	
BUILDING A HIGH PERFORMANCE MASS STORAGE SYSTEM FOR A WLCG TIER-1 SITE	204
V. Sapunenko, L. dell’Agnello, A. Cavalli, D. Gregori, A. Prosperini, P.P. Ricci, F. Noferini, E. Ronchieri, V. Vagnoni	
ATLAS DISTRIBUTED COMPUTING AUTOMATION	212
J. Schovancová, F.H. Barreiro Megino, C. Borrego, S. Campana, A. Di Girolamo, J. Elmsheuser, J. Hejbal, T. Kouba, F. Legger, E. Magradze, R. Medrano Llamas, G. Negri, L. Rinaldi, G. Sciacca, C. Serfon, D.C. Van Der Steron on behalf of the ATLAS Collaboration	
COMPUTING FACILITIES FOR SMALL PHYSICS ANALYSIS GROUP: EXAMPLES AND CONSIDERATION	216
A.Y. Shevel	
IMPLEMENTATION OF COMMON TECHNOLOGIES IN GRID MIDDLEWARES ..	220
O. Smirnova, B. Kónya, C. Aiftimiei, M. Cecchi, L. Field, P. Fuhrmann, J. K. Nilsen, J. White	
MATHCLOUD: FROM SOFTWARE TOOLKIT TO CLOUD PLATFORM FOR BUILDING COMPUTING SERVICES	228
O.V. Sukhoroslov	
DEPENDABLE JOB-FLOW DISPATCHING AND SCHEDULING IN VIRTUAL ORGANIZATIONS OF DISTRIBUTED COMPUTING ENVIRONMENTS	234
V.V. Toporkov, A.S. Tselishchev, D.M. Yemelyanov, A.V. Bobchenkov	

ADVANCEMENTS IN BIG DATA PROCESSING IN THE ATLAS AND CMS EXPERIMENTS	243
A.V. Vaniachine on behalf of the ATLAS and CMS Collaborations	
APPLICATION OF DATA GRID TECHNOLOGY FOR SHARING SCIENCE OUTREACH RESOURCES IN CHINA	249
Zhang Zuli, He Hongbo, Xiao Yun	
СОЗДАНИЕ В ОИИИ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ОБРАБОТКИ ДАННЫХ УРОВНЯ TIER-1 ЭКСПЕРИМЕНТА CMS НА LHC	254
Н.С. Астахов, С.Д. Белов, А.Г. Долбилов, В.Е. Жильцов, В.В. Кореньков, В.В. Мицын, Т.А. Стриж, Е.А. Тихоненко, В.В. Трофимов, С.В. Шматов	
ТЕСТИРОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ PVM И MPI С МИГРАЦИЕЙ ПРОЦЕССОВ MOSIX В РАСПРЕДЕЛЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ СРЕДЕ	266
А.В. Богданов, Е Мьинт Найнг, Пья Сон Ко Ко	
GRIDCOM, GRID COMMANDER: ГРАФИЧЕСКИЙ ИНТЕРФЕЙС ДЛЯ РАБОТЫ С ЗАДАЧАМИ И ДАННЫМИ В ГРИДЕ	270
В.В. Галактионов	
О ПРИМЕНЕНИИ ТЕХНОЛОГИИ CUDA ДЛЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ И РАСПОЗНАВАНИЮ ГРАФИЧЕСКИХ ОБРАЗОВ	274
И.М. Гостев	
МЕТОДЫ ВЫДЕЛЕНИЯ КЛАСТЕРОВ В БОЛЬШИХ СЕТЯХ ПЕРЕДАЧИ ДАННЫХ	280
Я.Р. Гринберг, И.И. Курочкин, А.В. Корх	
АЛГОРИТМЫ УВЕЛИЧЕНИЯ СУММАРНОГО ТРАФИКА В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ	286
Я.Р. Гринберг	
АРХИТЕКТУРА КОММУНИКАЦИОННОЙ СРЕДЫ СУПЕРКОМПЬЮТЕРОВ СЛЕДУЮЩЕГО ПОКОЛЕНИЯ И ТЕОРИЯ ПРОСТРАНСТВЕННО-ВЛОЖЕННЫХ СЛОЖНЫХ СЕТЕЙ	292
А.П. Демичев, В.А. Ильин, А.П. Крюков, С.П. Поляков	
ПАРАЛЛЕЛЬНЫЕ ТЕХНОЛОГИИ В ЗАДАЧЕ МАКСИМИЗАЦИИ ПРАВДОПОДОБИЯ	302
А.В. Ермилов	
РЕШЕНИЕ ПАРАЛЛЕЛЬНЫХ ЗАДАЧ ЗА РАМКАМИ КЛАССА EP В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СРЕДАХ	306
Ю.А. Жолудев	
ИССЛЕДОВАНИЕ ОСОБЕННОСТЕЙ ПРОБЛЕМЫ ИНТЕРОПЕРАБЕЛЬНОСТИ В GRID-ТЕХНОЛОГИИ И ТЕХНОЛОГИИ ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ	312
Е.Е. Журавлёв, В.Н. Корниенко, А.Я. Олейников	
ВОПРОСЫ ИНТЕРОПЕРАБЕЛЬНОСТИ В ОБЛАЧНЫХ ВЫЧИСЛЕНИЯХ	321
С.В. Иванов	

ПОИСК АССОЦИАТИВНЫХ ПРАВИЛ В БОЛЬШИХ МАССИВАХ ДАННЫХ С ПОМОЩЬЮ ГЕТЕРОГЕННОЙ ГРИД НА БАЗЕ VOINC	327
Е.Е. Ивашко, А.С. Головин	
ГРИДННС: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ	332
В.А. Ильин, В.В. Кореньков, А.П. Крюков	
СИСТЕМА МАССОВОЙ ИНТЕГРАЦИИ БАЗ ДАННЫХ: ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ И СПОСОБ РЕАЛИЗАЦИИ	337
В.Н. Коваленко, Е.И. Коваленко, А.Ю. Куликов	
МОДЕЛИРОВАНИЕ ГРИД СИСТЕМЫ OFF-LINE ОБРАБОТКИ ДАННЫХ ДЛЯ ЭКСПЕРИМЕНТА NISA	343
В.В. Кореньков, А.В. Нечаевский, В.В. Трофимов	
СОЗДАНИЕ ОБЛАЧНОЙ ПЛАТФОРМЫ УРОВНЯ Tier3 В ГРИД-ИНФРАСТРУКТУРЕ ЭКСПЕРИМЕНТОВ НА LHC ДЛЯ РАЗРАБОТКИ ПРИЛОЖЕНИЙ РАДИОЛОКАЦИОННОГО КОСМИЧЕСКОГО МОНИТОРИНГА	349
В.В. Кореньков, В.М. Котов, Н.А. Русакович, А.В. Яковлев	
ПОИСК РЕШЕНИЯ ВАРИАЦИОННОЙ ЗАДАЧИ В ВИДЕ МИНИМАЛЬНОГО ПУТИ НА ГРАФЕ	355
Д.Т. Лотарев	
ЭФФЕКТИВНЫЙ МЕТОД ПЛАНИРОВАНИЯ РЕСУРСОВ В ГЕТЕРОГЕННЫХ РАСПРЕДЕЛЕННЫХ СИСТЕМАХ И ЕГО РЕАЛИЗАЦИЯ В MAUI	359
С.В. Минухин, С.В. Баранник, С.В. Знахур, Р.И. Зубатюк	
КОНСОЛИДАЦИЯ ЭЛЕКТРОННЫХ БИБЛИОТЕЧНЫХ И ИНТЕРНЕТ-РЕСУРСОВ ДЛЯ ОБРАЗОВАТЕЛЬНЫХ И НАУЧНЫХ ЦЕЛЕЙ НА ОСНОВЕ GRID-ТЕХНОЛОГИЙ	365
Б.В. Олейников, А.И. Шалабай	
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ С ОТКРЫТЫМ ИСХОДНЫМ КОДОМ ДЛЯ ПОСТРОЕНИЯ И УПРАВЛЕНИЯ ОБЛАЧНЫМИ СРЕДАМИ НА РАСПРЕДЕЛЕННЫХ ГЕТЕРОГЕННЫХ ИНФРАСТРУКТУРАХ	371
А.В. Пярн	
ЭФФЕКТИВНЫЙ МОНИТОРИНГ КОММУНИКАЦИЙ НА ОСНОВЕ ВНЕШНЕЙ АППРОКСИМАЦИИ ГРАФА	377
А.М. Раппопорт	
ПАРАЛЛЕЛЬНЫЕ ЗАДАНИЯ В ГРИД-СРЕДЕ	383
М.М. Степанова, О.Л. Стесик	
ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ РАСПРЕДЕЛЕННОЙ ВИЗУАЛИЗАЦИИ НА ПРИМЕРЕ МВК НИЦ «КУРЧАТОВСКИЙ ИНСТИТУТ»	389
И.А. Ткаченко	
ПРИМЕНЕНИЕ ТЕХНОЛОГИИ ВИРТУАЛИЗАЦИИ ДЛЯ ИЗУЧЕНИЯ ПРИНЦИПОВ ФУНКЦИОНИРОВАНИЯ КОМБИНИРОВАННЫХ ВЫЧИСЛИТЕЛЬНЫХ ИНФРАСТРУКТУР	395
Н.П. Храпов	

ЭЛЕКТРОННАЯ БИБЛИОТЕКА НАУЧНОГО ЦЕНТРА	400
С.К. Шикота, С.А. Крашаков, Л.Н. Щур	
ПРОБЛЕМЫ РАЗВИТИЯ ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ ИНФРАСТРУКТУРЫ НАУЧНОГО ЦЕНТРА.....	405
Л.Н. Щур, С.К. Шикота	
INDEX	409

АРХИТЕКТУРА КОММУНИКАЦИОННОЙ СРЕДЫ СУПЕРКОМПЬЮТЕРОВ СЛЕДУЮЩЕГО ПОКОЛЕНИЯ И ТЕОРИЯ ПРОСТРАНСТВЕННО-ВЛОЖЕННЫХ СЛОЖНЫХ СЕТЕЙ¹

Демичев А.П.^{1,2}, Ильин В.А.^{1,2}, Крюков А.П.^{1,2}, Поляков С.П.²

¹Национальный исследовательский центр "Курчатовский институт"

Россия, 123182 Москва, пл. Академика Курчатова, д. 1

²Научно-исследовательский институт ядерной физики имени Д.В. Скобельцына

Россия, 119991, ГСП-1, Москва, Ленинские горы, д. 1, стр. 2

В работе предлагается подход к разработке коммуникационных сетей суперкомпьютеров следующего поколения. Рассмотрены алгоритмы построения сложных сетей со свойствами "малого мира", а именно, медленным (логарифмическим) ростом среднего расстояния между узлами при увеличении их числа. При этом сети, построенные на основе этих алгоритмов, имеют базовую структуру регулярной решетки с дополнительными перемычками между узлами, которые и обеспечивают свойства "малого мира". Поскольку вычислительные узлы располагаются в физическом трехмерном пространстве, и топологические аспекты сети коррелируют с пространственными аспектами, для анализа используются методы теории пространственно-вложенных сложных сетей.

We propose an approach to the interconnection network design for the next generation supercomputers. Algorithms of small-world complex networks construction with slow (logarithmic) growth of average distance between nodes are considered. The constructed networks have basic structure of the regular lattice with additional shortcuts which provide the small-world properties. Since the computing nodes are arranged in real three-dimensional space and topological aspects of the networks should correlate with spatial aspects, the methods of the theory of spatially embedded complex networks are used.

1 Введение

После того как был достигнут петафлопсный (10^{15} FLOPS) барьер производительности суперкомпьютеров, перед разработчиками высокопроизводительных вычислительных систем встал вопрос о принципах построения систем следующего поколения — с производительностью порядка эксафлопса (10^{18} FLOPS) [1]. Хотя появление реальных вычислительных систем такого уровня ожидается не ранее 2018–2020 года, подходы и принципы их построения начинают интенсивно разрабатываться уже сейчас, поскольку на пути к их построению предстоит решить ряд сложных научно-технических задач и выработать принципиально новые решения для их архитектуры и аппаратной реализации.

Одной из важнейших составляющих любого суперкомпьютера является коммуникационная сеть, которая в первую очередь определяет возможность увеличения числа вычислительных узлов, что необходимо для достижения желаемой производительности. Как ожидается (см., например, [2]), суперкомпьютеры эксафлопного уровня будут иметь порядка 100 000 и более вычислительных узлов. Таким образом, одной из важнейших задач, которую предстоит решить на пути к построению суперкомпьютеров следующего поколения, является разработка коммуникационных сетей с хорошими свойствами масштабируемости и возможностью эффективно обслуживать огромное число вычислительных узлов.

Тремя основными аспектами проектирования коммуникационных сетей, которые в наибольшей степени определяют их функциональные свойства, являются:

- топология сети (network topology);

¹ Работа частично финансируется РФФИ, грант 12-07-00408-а.

- метод управления потоками (flow control); иногда используют термин "метод переключения (switching method)";
- алгоритм маршрутизации (routing algorithm).

В данной работе обсуждается, в основном, именно топология коммуникационной сети (в том смысле, в котором термин "топология" используется в теории сетей). Два других аспекта очень важны, но выходят за рамки текущего обсуждения. Выбор подходящей топологии жизненно важен для проектирования сети, поскольку маршрутизация и механизмы управления потоком в большой степени основаны на ее свойствах.

В идеальном случае коммуникационная сеть должна была бы быть полностью соединена (полный граф), чтобы позволить одновременную непосредственную связь между всеми парами узлов, достигая оптимальной пропускной способности и задержки. Этот подход может быть применен к системам с немногими узлами, но он не масштабируется на большие сети, так как число связей для каждого узла было бы равно числу всех узлов сети минус единица. Пропускная способность сети должна масштабироваться с ростом числа процессоров, что обеспечивается правильной комбинацией хорошего выбора топологии и алгоритмов маршрутизации.

Существуют два общих типа сетей:

- прямые сети (direct networks), в которых каждый узел является терминальным, действуя и как источник, и как приемник для сообщений, а также и как рутер для управления входящими сообщениями;
- не прямые сети (indirect networks) содержат "нетерминальные" узлы (рутеры), которые используются только для маршрутизации.

Не прямые сети имеют свои достоинства для ограниченного числа узлов, но плохо масштабируются. Поэтому мы будем рассматривать только прямые сети. Более точно, мы будем рассматривать обобщения регулярных решеток с топологией D -мерных торов. В литературе, посвященной сетям, для такой топологии часто используется термин " k -ary n -cube" [3] (n — размерность тора, в наших обозначениях $n = D$). При этом каждый узел, как в любой прямой сети, является рутером. Известно, что такие сети при большом числе узлов имеют преимущества по сравнению с другими архитектурами, например, гиперкубами высоких размерностей (см., например, [4]).

Важным аргументом в пользу использования регулярных решеток является тот факт, что на такую структуру коммуникационной сети естественным образом отображаются параллельные вычислительные задания, связанные с численным моделированием D -мерных объектов. В частности, коммуникационные сети со структурой трехмерной решетки оптимальны для моделирования трехмерных реальных объектов, а именно такого типа задачи, как предполагается, будут составлять значительную долю задач, решаемых на суперкомпьютерах экзафлопного уровня.

Однако при огромном числе узлов, характерном для компьютеров следующего поколения, архитектура регулярных решеток с топологией D -мерных торов имеет и существенные недостатки. В частности, решетки невысокой размерности имеют весьма большую среднюю длину пути между узлами, а решетки высокой размерности, сравнимой с логарифмом числа узлов, трудно реализовать технически из-за большой длины физических коммуникационных каналов. С другой стороны, известно, что наилучшими структурами вычислительных систем по различным критериям функционирования, например, производительности и надежности, при одинаковом числе вычислительных узлов и каналов связи являются структуры с минимальным средним расстоянием между узлами (см., например, [5]). Поэтому обычные сети с простой структурой регулярных решеток окажутся недостаточно эффективными для решения задач более общего типа, не связанных с триангуляцией трехмерных объектов.

В связи с этим представляется перспективным использовать для построения коммуникативных сетей для экзафлопсных компьютеров сети со свойствами "малого мира" [6], одним из важнейших свойств которых является малое среднее расстояние между узлами и малый диаметр сети. Более точное выражение этого свойства заключается в следующем: для

регулярной D -мерной решетки среднее расстояние d между узлами растет как степень числа узлов: $d \sim N^{1/D}$, а для сети со свойствами малого мира существенно медленнее: $d \sim \ln N$. Заметим, что в некоторых сложных сетях среднее расстояние зависит от числа узлов степенным образом, как и в случае регулярных решеток, однако показатель степени существенно меньше:

$d \sim N^\gamma$, $\gamma \ll 1/D$. Зачастую исследования свойств сложных сетей осуществляются с помощью численного моделирования, и отличить такое степенное поведение от логарифмического весьма непросто. С точки зрения использования таких сетей в прикладных целях при достаточно малом показателе отличие тоже является несущественным. Поэтому можно считать, что такие сети также обладают свойствами малого мира (в широком смысле).

В классическом варианте [6] сложные сети со свойствами малого мира получаются на промежуточной стадии в процессе стохастической трансформации регулярных решеток в полностью случайные графы Эрдеша-Реньи [7], [8]. При этом структура регулярной решетки нарушается, что, как отмечалось выше, нежелательно для коммуникационных сетей экзафлопсных компьютеров. Поэтому в данной работе предлагается использовать модификацию способа построения сетей с малой средней длиной пути между узлами, при которой сохраняется базовая решеточная структура, но к ней определенным образом добавляются дополнительные связи, называемые перемычками, которые и обеспечивают свойства малого мира. Схематический вид таких сетей в одномерном и двумерном случаях представлен на рис.1. Для краткости мы в дальнейшем будем использовать для таких сетей термин "решеточные сети с перемычками" (РСП).

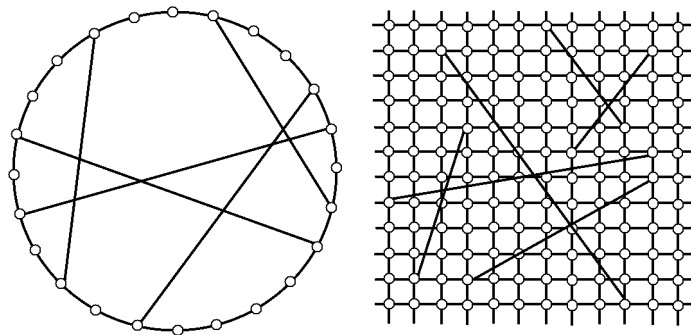


Рисунок 1: Схематический вид решеточных сетей с добавленными перемычками в одномерном (слева) и двумерном (справа) случаях

Необходимо отметить, что помимо малой средней длины пути между узлами, еще одним общим отличительным свойством сетей со свойствами малого мира является высокая степень кластеризации [6], [8]. Высокая кластеризация обеспечивает локальную устойчивость сети: существование локальных обходных путей при выходе из строя какого-либо узла сети. Однако в нашем случае такую локальную устойчивость (существование локальных обходных путей) для $D > 1$ обеспечивает решеточная основа, и поэтому мы не будем обсуждать это свойство малого мира в данной работе.

Длина пути (расстояние) между узлами понимается в сетевом смысле: как минимальное число ребер, по которым надо пройти, чтобы попасть из одного узла в другой. Соответственно среднее расстояние между узлами определяется как среднее по всем парам узлов данной сети. Однако для больших сетей, какой ожидается коммуникационная сеть экзафлопсных компьютеров, определенная таким образом длина пути между узлами может оказаться неадекватной характеристикой, поскольку для нахождения кратчайших маршрутов необходимо знать глобальную структуру сети. Соответственно маршрутизация сообщений, использующая кратчайшие пути, может оказаться слишком сложной и неэффективной, так как связана с хранением и обработкой большого объема информации. Поэтому особую важность приобретают алгоритмы маршрутизации, основанные на локальной навигации [9], [10]. Задача локальной навигации в сетях ставится следующим образом: сообщение передается от узла к узлу по ребрам (связям); узел "знает" географическое положение (другими словами, положение

в базовой решетке) всех узлов и своих *ближайших* сетевых соседей с учетом перемычек (может знать соседей на глубину больше единицы; но информация обо всех перемычках в сети не используется); необходимо доставить сообщение в узел назначения по возможно кратчайшему пути. В простейшем варианте эту задачу решает так называемый жадный алгоритм (англ. greedy algorithm; иногда называется также алгоритмом экономного продвижения): текущий узел пересылает сообщение тому из своих соседей, который географически (то есть в смысле координат на решетке) ближе всего к цели (узлу назначения).

Таким образом, в данной работе мы исследуем среднюю глобальную и среднюю навигационную длины пути между узлами сети, как важнейшие характеристики, определяющие коммуникационные свойства сети. Основной целью работы является разработка оптимального алгоритма построения сети с большим числом узлов, но малой средней глобальной или навигационной длиной пути между узлами. Общая идея состоит в добавлении к решеточной основе дополнительных перемычек по специальному алгоритму (или алгоритмам), так, чтобы оптимизировать соотношение "цены" и "качества" для получаемой таким образом сети. В качестве "цены" выступает удельная длина дополнительных перемычек (общая длина перемычек в единицах базовой решетки, деленная на число узлов сети), а "качество" — это глобальная или навигационная средняя длина пути между узлами.

2 Оптимизация параметров коммуникационной сети

Как уже упоминалось, оригинальный алгоритм получения сложной сети со свойствами малого мира [6] является стохастическим: на каждом шаге алгоритма ребра графа меняют свое положение с некоторой вероятностью. В результате многократного применения такого алгоритма возникает ансамбль графов с некоторым распределением их характеристик, в частности, с некоторым распределением средней длины пути между узлами экземпляра графа. Для многих реальных сетей стохастический процесс их образования оказывается внутренне присущим (так, это справедливо для сети Интернет; другие примеры стохастических сетей см., например, в [8]). Проектирование коммуникационной сети суперкомпьютера находится под контролем разработчика, и стохастичность не является внутренне присущим элементом этого процесса. Существует ряд работ [11] – [15], в которых предложены детерминистские алгоритмы построения сетей со свойствами малого мира. Предварительное исследование этих сетей показывает, что по соотношению "цены" и "качества" они проигрывают стохастическим сетям, особенно если показателем качества является навигационная длина. В любом случае свойства сетей, полученных с помощью детерминистских алгоритмов, должны быть сопоставлены с лучшими экземплярами сетей, которые можно получить с помощью стохастических алгоритмов и компьютерного моделирования. В данной работе мы ограничимся исследованием стохастических алгоритмов и разработкой методологии сравнения различных сетей. Кроме того, хотя часть полученных результатов справедлива для решеток произвольной размерности, в данной работе рассматривается простейший случай одномерной решетки с топологией окружности как на рис. 1 слева. Основными причинами для этого являются:

- удобство отработки различных алгоритмов и сравнения свойств получающихся сетей (сокращение времени численного моделирования);
- исследования свойств одномерных сетей могут представлять непосредственный прикладной интерес: например, при использовании так называемой DOR-маршрутизации (Dimension Ordered Routing; см., например, [16]), при котором сообщения двигаются вдоль каждого измерения независимо от других измерений, и это движение определяется свойствами одномерной (модифицированной) решетки.

Обобщения на решетки более высоких размерностей и результаты сравнения с сетями, получаемыми с помощью детерминистских алгоритмов, будут представлены в последующих работах.

Стохастический алгоритм построения сетей малого мира с сохранением базовой решетки предложен в ряде работ, в частности, в работах [17] (см. также обзор [18] и ссылки в нем). В одномерном случае он формулируется следующим образом.

Алгоритм S1: (1) исходным объектом является одномерная решетка с L узлами и

топологией окружности; (2) последовательно перебираются все узлы решетки и с вероятностью $0 < p \leq 1$ к каждому узлу i подсоединяют первый конец перемычки; (3) второй конец перемычки (то есть, узел решетки j , в который она входит) не может совпадать с соседями исходного узла в смысле базовой решетки и приводит к дублированию уже существующей перемычки, а в остальном выбирается случайно с вероятностью $P(r) \sim r^{-\alpha}$, которая является степенной функцией решеточного расстояния $r=r_{ij}$ между узлами.

Зависимость вероятности перемычки от расстояния между узлами отражает корреляцию между топологическими и пространственными свойствами сети [18].

При использовании алгоритма S1 построения PCП управляющими параметрами или, другими словами, параметрами, характеризующими ансамбль (аналог температуры для обычного канонического ансамбля в статистической физике), являются параметры L, p, α . Соответственно, эти параметры должны быть оптимизированы при построении сети на основе этого алгоритма.

При исследовании возможной архитектуры коммуникационной сети (которой внутренне не присуща стохастичность появления перемычек) может быть удобно использовать следующую модификацию алгоритма S1.

Алгоритм S1m: (1) исходным объектом является одномерная решетка с L узлами и топологией окружности (как в S1); (2) фиксируется число перемычек t , которые должны быть добавлены к решетке; (3) из всех L узлов решетки случайным образом выбираются t узлов, к которым подсоединяются первые концы перемычек; (4) второй конец каждой из t перемычек выбирается случайно и полностью идентично шагу 3 алгоритма S1.

При использовании алгоритма S1m построения PCП управляющими параметрами являются параметры L, t, α . Отличие от базового алгоритма S1 заключается в том, что в если в базовом алгоритме случайным является как расположение, так и число перемычек, в алгоритме S1m случайным является только положение перемычек. Другими словами, степень стохастичности алгоритма уменьшается за счет сужения вероятностного пространства событий (конкретных реализаций PCП). С другой стороны, объединение вероятностных пространств алгоритма при всех допустимых значениях t эквивалентно объединению пространств событий базового алгоритма при допустимых значениях параметра p . Поэтому поиск оптимальной PCП, построенной по алгоритмам S1 и S1m, при достаточно репрезентативной выборке должен приводить к одинаковым результатам.

Можно попытаться еще уменьшить степень стохастичности алгоритма построения PCП следующим образом.

Алгоритм S2: (1) исходным объектом является одномерная решетка с L узлами и топологией окружности (как в S1 и S1m); (2) фиксируется общее число перемычек t , которые должны быть добавлены к решетке (как в алгоритме S1m) и фиксируется число $c < t$ перемычек, которые будут добавлены специальным образом; (3) из всех L узлов решетки случайным образом выбираются $t-c$ узлов, к которым подсоединяются первые концы перемычек; (4) второй конец каждой из $t-c$ перемычек выбирается случайно и полностью идентично шагу 3 алгоритма S1; (5) c перемычек добавляются специальным образом; перемычки добавляются последовательно одна за другой, при добавлении каждой перемычки: составляется список узлов, к которым уже присоединены перемычки (как перемычки, построенные на предыдущих шагах 2–4, так и новые, уже построенные на этом шаге перемычки); из этого списка узлов случайным равновероятным образом выбирается один и к нему подсоединяется новая перемычка; второй конец дополнительных перемычек выбирается так же как на шаге 4 (при этом новая перемычка не должна совпадать с предыдущими и с ребром базовой решетки).

При использовании алгоритма S2 построения PCП управляющими параметрами или, другими словами, параметрами, характеризующими ансамбль, являются параметры L, t, c, α . В алгоритме S2 не только число перемычек детерминировано, но положение их не вполне случайно, а подчиняется некоторым дополнительным правилам, причем степень этой регулярности определяется параметром c (при $c = 0$ алгоритм S2 совпадает с S1m). Как

показывает численное моделирование, среднее расстояние между узлами РСП S2 в широком диапазоне значений параметра c меньше, чем среднее расстояние в РСП, построенных по алгоритмам S1 и S1m.

Среднее сетевое расстояние d между узлами получается после двойного усреднения: по стохастическому процессу создания перемычек (другими словами, по статистическому ансамблю случайных графов, полученному в результате стохастического процесса образования перемычек) и по всем парам узлов графа. Характерное поведение средней длины в зависимости от параметров ансамбля из 1000 РСП, построенного с помощью алгоритма S1m, полученное нами с помощью численного моделирования для сети, состоящей из $L = 10000$ узлов, показано на рис. 2.

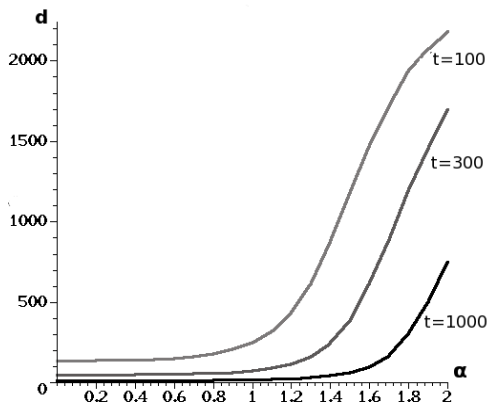


Рисунок 2: Средняя длина в зависимости от параметров ансамбля РСП (алгоритм S1m)

рассматривать зависимость среднего расстояния от этой величины. Характерные результаты численного моделирования для такой зависимости приведены на рис. 3.

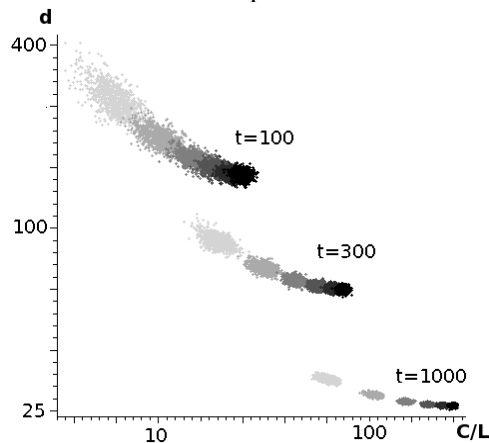


Рисунок 3: Зависимость распределений средних по экземплярам расстояний d между узлами от удельной длины перемычек

При $\alpha > 2$ средняя длина перемычек оказывается слишком малой и они не оказывают существенного влияния на свойства сети, так что свойства малого мира исчезают [17]. При наивном рассмотрении результатов, представленных на рис. 2 (для алгоритма S1 и параметра p зависимость аналогична), можно прийти к заключению, что для получения сети с наилучшей характеристикой расстояния между узлами надо просто выбирать наибольшее значение параметра t (или p) и наименьшее значение α . Однако, как отмечалось выше, каждое решение имеет свою "цену" и необходимо оптимизировать соотношение "цены" и "качества" для получаемой сети (иначе очевидным решением является просто полный граф, в котором все узлы соединены прямыми связями). Поскольку в данной работе в качестве "цены" выступает удельная длина дополнительных перемычек, необходимо

рассматривать зависимость среднего расстояния от этой величины. Каждая точка с соответствующей градацией серого цвета на рис. 3 соответствует экземпляру РСП в ансамбле со значениями $\alpha=0, 0.2, 0.4, 0.6, 0.8, 1.0$ (чем больше значение α , тем светлее оттенок серого цвета) и $t = 100, 300, 1000$, ось ординат соответствует среднему расстоянию между узлами, а по оси абсцисс отложена реальная удельная длина перемычек C/L (C — общая длина перемычек в единицах базовой решетки; L — число узлов сети), полученная в результате численного моделирования для данного экземпляра ансамбля (по обеим осям использован логарифмический масштаб). Для алгоритма S1 получены аналогичные результаты, но с более широким, как и ожидалось, распределением расстояний по

ансамблю. Из представленных данных видно, что при одинаковой цене для уменьшения средней длины надо выбирать ансамбль с возможно большим значением t (или p), и, одновременно, с наибольшим возможным значением α (пока это возможно для данной удельной длины перемычек). Другими словами, большое количество не слишком длинных перемычек является более дешевым решением. Это контрастирует с результатами для простой зависимости средней длины от параметров p и α , которые при "наивном" рассмотрении могут привести к заключению, что надо просто выбирать наименьшее значение α .

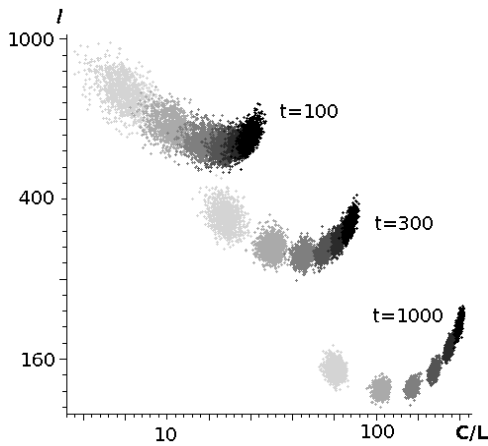


Рисунок 4: Зависимость распределений средних по экземплярам навигационных длин l от удельной длины перемычек

Как отмечено во Введении, с точки зрения оптимизации параметров стохастических РСП представляет интерес выяснение зависимости навигационной длины от параметров алгоритмов. На рис. 4 показаны распределения средней (по экземплярам) навигационной длины l в зависимости от удельной длины перемычек и параметра α при различных значениях t для алгоритма S1m. В случае алгоритма S1 поведение оказывается аналогичным при замене t на соответствующее значение параметра p . Видно, что при фиксированном числе перемычек t существует минимум l , достигаемый при различных значениях α , зависящих от t . Как и можно было ожидать, чем большее число перемычек добавляется к

решетке, тем меньшая средняя длина получается в минимуме по α . Видно также, что для заданной "цены" можно подобрать параметры алгоритма S1m с наименьшей навигационной длиной, или, наоборот, для заданной навигационной длины можно подобрать параметры алгоритма S1m с наименьшей удельной длиной перемычек. Поэтому и в этом случае надо решать оптимизационную задачу "цена – качество".

Представленные детальные данные показывают, что использование только зависимостей качества от цены (средней длины от удельной длины перемычек) не позволяет однозначно выбрать оптимальную сеть, то есть, в случае алгоритмов S1 или S1m, выбрать оптимальные параметры p , t , α . Поэтому надо использовать методы так называемой многокритериальной оптимизации.

Достаточно часто в реальных ситуациях качество эксплуатации исследуемого объекта или системы оценивается не единственным критерием или показателем качества, а совокупностью таких критериев. Такая постановка задачи приводит к задаче оптимизации с векторной целевой функцией, которая должна трактоваться неким определенным образом. В нашем случае мы будем учитывать два показателя: среднюю длину пути между узлами (глобальную d или навигационную l), что выражает "качество" РСП, и удельную длину перемычек C/L , что отражает "цену" построения РСП. Очевидно, что эти две величины взаимосвязаны: увеличивая цену C/L можно улучшить качество (уменьшить d или l), и, наоборот, улучшение качества зачастую связано с увеличением цены. Существует ряд подходов и методов для решения задач такой многокритериальной оптимизации (см., например, [19]). Мы будем использовать один из простейших и наглядных методов, а именно, метод взвешенных сумм (в более общем контексте такой подход называется скаляризацией многокритериальной оптимизации). Для этого введем следующие скалярные целевые функции $G_w = w d + (1-w) C/L$ и $G'_w = w l + (1-w) C/L$. Минимизация этих целевых функций означает, что подобраны оптимальные значения параметров алгоритмов с точки зрения качества (малой длины пути между узлами) и цены (малой длины перемычек). При этом параметр $0 \leq w \leq 1$ характеризует относительную значимость каждого из критериев (качество и цена). Другими словами, предлагаемый способ оптимизации предполагает, что для каждого значения параметра значимости критериев w должны быть подобраны значения параметров алгоритмов (например, p/t и α для S1/S1m; t , c и α для S2; размер сети L считаем заданным), которые минимизируют G_w или G'_w .

Для РСП, построенных по алгоритмам S1m с числом узлов $L=10000$, численно найдены параметры ансамблей t и α , при которых достигаются минимальные значения целевых функций при различных значениях значимости качества w . На рис. 5 эти данные представлены в случае G_w для параметра α , а на рис. 6 — для параметра t .

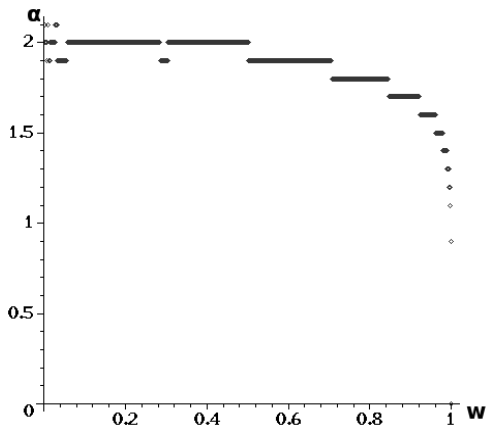


Рисунок 5: Зависимость оптимального значения α (минимизирующего G_w) от параметра значимости качества w

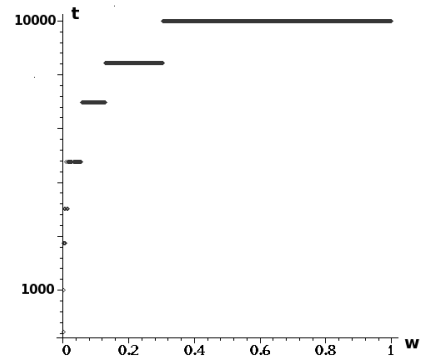


Рисунок 6: Зависимость оптимального значения t (минимизирующего G_w) от параметра значимости качества w

Из этих результатов следует, что когда значимость качества w превышает значение ~ 0.35 , число перемычек должно быть максимально возможным (равным числу узлов сети); при меньших значениях этого параметра важность минимизации длины перемычек приводит к уменьшению числа перемычек (значения t). Напротив, значения параметра α близкие к 2 (то есть к критическому значению с точки зрения сохранения свойств малого мира [17]) является оптимальным при значениях параметра w меньших значения ~ 0.5 , то есть когда доминирует цена.

Как отмечалось выше, при локальной навигации можно использовать информацию о сетевых соседях на глубину больше единицы. В частности, можно рассмотреть вариант локальной навигации, когда просматриваются не только ближайшие соседи, но и соседи соседей. При этом сообщение на следующем шаге пересылается в тот соседний узел, один из соседей которого ближе всего к узлу назначения в смысле решеточной метрики. Хотя при такой двухуровневой навигации на каждом шаге объем вычислений несколько увеличивается по сравнению с обычным жадным алгоритмом, но алгоритм остается локальным (не вычисляется весь путь до адресата, и объем не зависит от размеров системы). Поэтому этот алгоритм является хорошо масштабируемым и приемлем для сверхбольших коммуникационных сетей. Для такой навигации можно определить соответствующую целевую функцию $G''_w = w l^{(2)} + (1 - w) C/L$, где $l^{(2)}$ — навигационная длина при двухуровневой навигации.

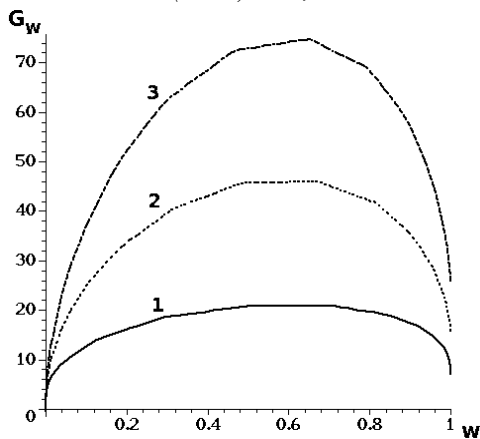


Рисунок 7: Зависимость минимальных значений целевых функций от параметра значимости w для алгоритма S1m: кривая 1 - $\min G_w$; кривая 2 - $\min G''_w$; кривая 3 - $\min G'_w$

На рис. 7 представлены результаты вычисления минимальных значений целевых функций G_w , G'_w и G''_w для сетей с числом узлов $L = 10000$, построенных по алгоритму S1m (усреднение по ансамблям, состоящим из 100 – 1000 экземпляров). Каждая точка кривых на рис. 7 соответствует ансамблю сетей с параметрами t , α , обеспечивающими минимум G_w , G'_w или G''_w при данном значении параметра значимости критерия качества w . Видно, что во всем диапазоне w наименьшие значения имеет функция G_w ("бесконечная" глубина просмотра при

навигации), минимальные значения G''_w (глубина просмотра — два) превышают G_w , а значения G'_w (единичная глубина просмотра) являются самими большими. Этот результат является интуитивно понятным: чем большая информация используется о структуре сети (большая глубина просмотра), тем меньше взвешенная сумма соответствующей средней

длины пути и удельной длины перемычек.

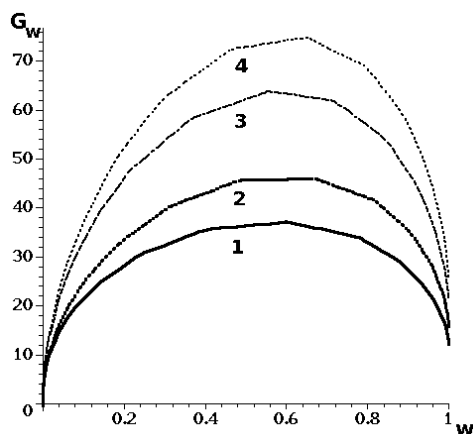


Рисунок 8: Зависимость минимальных значений целевых функций от параметра значимости w : 1 - $\min G''w$ для алгоритма S2; 2 - $\min G''w$ для алгоритма S1m; 3 - $\min G'w$ для алгоритма S2; 4 - $\min G'w$ для алгоритма S1m

Более важно, что целевые функции G_w , G'_w и G''_w позволяют сравнивать различные алгоритмы построения РСП, в том числе с различным набором параметров. В частности, результаты сравнения для алгоритмов S1m и S2 представлены на рис. 8.

Результаты сравнения минимальных значений целевых функций показывают, что при использовании маршрутизации сообщений, основанной как на локальной навигации с двухуровневой глубиной просмотра, так и на одноуровневом жадном алгоритме, для построения сетей более предпочтительным является алгоритм S2.

3 Заключение

В работе предложен подход к разработке коммуникационных сетей суперкомпьютеров следующего поколения, которые обладают свойствами "малого мира", а именно, медленным (логарифмическим) ростом среднего расстояния между узлами при увеличении их числа. При этом рассмотренные сети имеют базовую структуру регулярной решетки, что является важным для распараллеливания вычислительных заданий, связанных с численным моделированием D -мерных объектов. Эта базовая структура дополняется перемычками между узлами, которые и обеспечивают свойства "малого мира". Предложены методы оптимизации соотношения "цены" и "качества" для получаемой таким образом сети, причем в качестве "цены" выступает удельная длина дополнительных перемычек (общая длина перемычек в единицах базовой решетки, деленная на число узлов сети), а "качество" — это глобальная или навигационная средняя длина пути между узлами. Эти методы также позволяют количественно сравнивать сети, полученные с помощью различных алгоритмов их построения.

В последующих публикациях будут исследованы сети со свойствами малого мира, построенные на основе детерминистских алгоритмов, а также продолжено рассмотрение свойств предложенных в данной работе сетей, как с точки зрения других характеристик (в частности, нагрузки на узлы, устойчивости и т.п.), так и обобщения на более высокие размерности базовой решетки.

Литература

- [1] Shainer G., Sparks B., Graham R. Toward Exascale computing, HPC Advisory Council, http://www.hpcadvisorycouncil.com/pdf/Toward_Exascale_computing.pdf;
Концепция создания эксафлопсного суперкомпьютера России: "Эксафлопные технологии. Концепция по развитию технологии высокопроизводительных вычислений на базе суперэвм эксафлопного класса (2012–2020 гг.)", http://filearchive.cnews.ru/doc/2012/03/esk_tex.pdf
- [2] Report on Institute for Advanced Architectures and Algorithms Interconnection Networks Workshop 2008, Future Technologies Group Technical Report Series, Oak Ridge, Tennessee USA; <http://www.csm.ornl.gov/workshops/IAA-IC-Workshop-08>
- [3] Dally W. J., Towles B. P. Principles and Practices of Interconnection Networks.- Amsterdam: Elsevier Science, 2003.- 550 p.
- [4] Dally W. J. Performance Analysis of k-ary n-cube Interconnection Networks, IEEE Transactions

- on Computers **39** (1990) 775.
- [5] Kleinrock L. *Communication Nets: Stochastic Message Flow and Design*.- New York: McGraw-Hill, 1964.- 220 p.
- [6] Watts D. J., Strogatz D. H. Collective dynamics of small-world networks, *Nature* **393** (1998) 440.
- [7] Erdős P., Rényi A., On the evolution of random graphs, *Publ. of the Math. Inst. of the Hungarian Academy of Sciences* **5** (1960) 17.
- [8] Albert R., Barabasi A.-L. Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74** (2002) 47.
- [9] Milgram S., The small world problem, *Psychology Today*, **2** (1967) 60.
- [10] Kleinberg J. M. Navigation in the small world, *Nature* **406** (2000) 845.
- [11] Zou Zhi-Yun et al., Regular Small-World Network, *Chin. Phys. Lett.* **26** (2009) 110502.
- [12] Boettcher S. , Goncalves B., Azaret J., Geometry and Dynamics for Hierarchical Regular Networks, *Journal of Physics A* **41** (2008) 335003.
- [13] Boettcher S., Goncalves B., Guclu H., Hierarchical Regular Small-World Networks, *J. Phys. A* **41** (2008) 252001.
- [14] Comellas F., Ozona J., Peters J. G., Deterministic small-world communication networks, *Information Processing Letters* **76** (2000) 83;
Comellas F. , Mitjana M., Peters J.G., Broadcasting in Small-World Communication Networks, In: *Proc. 9th Int. Coll. on Structural Information and Communication Complexity* (2002), eds. C. Kaklamanis and L. Kirousis, pp. 73–85.
- [15] Xiao W., Parhami B., Cayley graphs as models of deterministic small-world networks, *Information Processing Letters* **97** (2006) 115.
- [16] Dally W.J. , Seitz C., Deadlock-free message routing in multiprocessor interconnection networks, *IEEE Trans. Comput.* **36** (5) (1987).
- [17] Moukarzel C.F. , de Menezes M. A., Shortest paths on systems with power-law distributed long-range connections, *Phys. Rev. E*, **65** (2002) 056709;
Sen P., Chakrabarti B., Small-world phenomena and the statistics of linear polymer, *J. Phys. A* **34** (2001) 7749;
Petermann T., De Los Rios P., Spatial small-world networks: A wiring-cost perspective (2005) arXiv:cond-mat/0501420;
Petermann T., De Los Rios P., Physical realizability of small-world networks *Phys. Rev. E* **73** (2006) 026114.
- [18] Barthelemy M., *Spatial Networks*, *Phys. Reports* **499** (2011) 1.
- [19] Steuer, R.E. *Multiple Criteria Optimization: Theory, Computations, and Application*.- New York: John Wiley and Sons, 1986.- 330 p.