УДК 004.738

ОПТИМИЗАЦИЯ РАСПРЕДЕЛЕНИЯ ЗАПРОСОВ МЕЖДУ КЛАСТЕРАМИ ОТКАЗОУСТОЙЧИВОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ

В.А. Богатырев, А.В. Богатырев, И.Ю. Голубев, С.В. Богатырев

Предложена оценка надежности распределенных вычислительных систем, предусматривающих перераспределение запросов при изменениях потоков запросов, отказах и отключениях узлов системы, объединяемых в совокупность кластеров. Предложена и решена задача оптимизации процесса перераспределения запросов между кластерами с учетом его влияния на задержки обслуживания и надежность системы.

Ключевые слова: оптимизация, надежность, перераспределение запросов, кластер, отказоустойчивость.

Введение

Повышение отказоустойчивости, надежности и производительности распределенных вычислительных систем, объединяющих в единую систему множество отдельных кластеров [1–3], достигается в результате динамического перераспределения запросов [4–7] между ними с учетом изменений загруженности кластеров, отказов и временных отключений их узлов.

В распределенной инфраструктуре [1–3], консолидирующей множество ресурсов, объединенных в кластеры, перераспределение запросов (нагрузки) может осуществляться между узлами как одного, так и различных кластеров, соединенных через сеть. При перераспределении запросов между кластерами увеличиваются издержки на взаимосвязь через сеть, но возрастают возможности балансировки загрузки и адаптации к отказам и отключениям узлов, что обусловливает актуальность оптимизации процесса распределения запросов [8, 9].

Задача оптимизации системы

Объектом исследования является распределенная вычислительная система (рис. 1), включающая M локальных кластеров и общедоступный кластер, объединяющий m серверов.

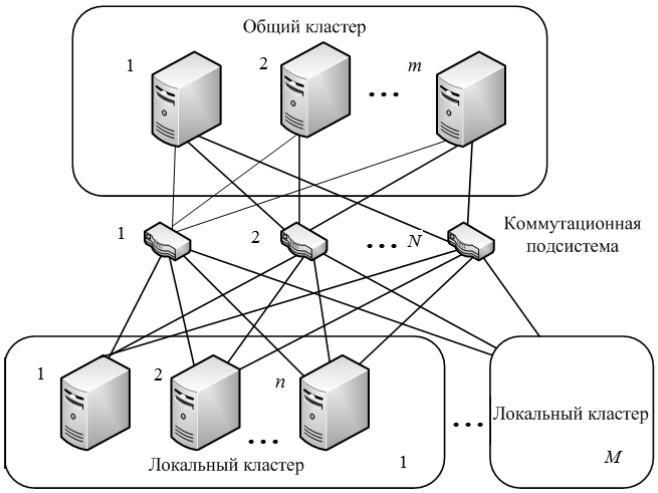

Рис. 1. Структура распределенной системы

В результате перераспределения запросов от локальных кластеров в общедоступный кластер обеспечивается сбалансированность нагрузки узлов системы и устойчивость системы к отказам и перегрузкам серверов локальных кластеров. Перераспределение запросов от некоторого локального кластера, содержащего в исходном (до отказов) состоянии n серверов, в общедоступный кластер осуществляется через N резервированных коммутационных узлов (маршрутизаторов или коммутаторов) [9].

При оптимизации структуры определяется число (кратность резервирования) серверов в локальных кластерах n и в общем кластере m, а также число коммутационных узлов N, обеспечивающие наибольшую надежность системы P при заданных ограничениях на стоимость построения системы s. При оценке надежности системы, в отличие от [9], где условие работоспособности системы сформулировано как требование сохранения в каждой подсистеме хотя бы одного узла, в предлагаемой работе учитываются нижние ограничения на число узлов в подсистемах, при которых не возникают перегрузки соответствующих кластеров.

При оптимизации процесса распределения запросов с учетом возможности отказов и отключений узлов общедоступного кластера будем считать заданными средние времена выполнения запросов в серверах кластеров и в коммутационных узлах v_0, v_1, их интенсивности отказов λ_0, λ_1, и восстановлений μ_0, μ_1. Будем считать известными вероятности r нахождения во включенном состоянии серверов общедоступного кластера. Оптимизация проводится при заданной интенсивности потока запросов λ, поступающего в локальный кластер и при необходимости перераспределяемого через сеть в общедоступный кластер, на который от других кластеров системы через сеть дополнительно направляется поток запросов с интенсивностью $\Lambda = \beta\lambda$.

При оптимизации структуры будем считать стоимости серверов локальных и общедоступного кластера, а также стоимость коммутационных узлов соответственно равными c_0, c_1, c_2.

В результате оптимизации процесса распределения потока запросов, поступающего в локальный кластер, ищется их доля, перераспределяемая через сеть в общедоступный кластер, при которой минимизируется среднее время пребывания запросов T.

Отличие предлагаемой задачи оптимизации распределения запросов от [9] заключается в учете возможностей отказов, восстановлений и отключений серверов общедоступного кластера в процессе функционирования. Учет возможности отключения серверов общедоступного кластера обусловлен тем, что предоставляемые им услуги по обслуживанию внешних для него запросов могут проводиться в фоновом режиме и поэтому могут отбрасываться при высокой нагрузке серверов, при решении важных для владельца кластера (сервера) задач, при профилактическом обслуживании или временных отключениях узлов по другим причинам.

Оценка надежности системы

Определим вероятность работоспособности системы для локального кластера из n серверов с учетом возможности использования в качестве резерва ресурсов m серверов общедоступного кластера, связь с которым обеспечивается через N коммутационных узлов.

Предположим, что пропускная способность каждого коммутационного узла достаточна, чтобы не ограничивать возможности перераспределения запросов, т.е. если исправен хотя бы один коммутационный узел, то запросы могут перераспределяться в общедоступный кластер, но для реализации такого пе-

перераспределения в локальном кластере должен быть исправен хотя бы один вычислительный узел. С учетом этих условий вероятность работоспособности системы составляет

$$P = (1-P_1)\left[\sum_{i=a}^{n} C_n^i p_0^i (1-p_0)^{n-i}\right] + P_1 \sum_{j=b}^{n+m} \left(C_{m+n}^j - d_j C_m^j\right) p_0^j (1-p_0)^{n+m-j}, \quad (1)$$

где $d_j = 1$, если $j \le m$, иначе $j = 0$; $P_1 = \sum_{i=1}^{N} C_N^i p_1^i (1-p_1)^{N-i}$ – вероятность исправности коммутационной подсистемы, при этом из соображений отсутствия перегрузки кластеров значения a и b определяются как ближайшие целые, бóльшие λv_0 и $\lambda(1+\beta)v_0$.

Надежность узлов определим по коэффициентам готовности, вычисляемым для серверов и коммутационных узлов соответственно как [10, 11]

$p_0 = \mu_0 / (\lambda_0 + \mu_0)$; $p_1 = \mu_1 / (\lambda_1 + \mu_1)$.

Формула (1) не учитывает возможность случайных временных отключений серверов общедоступного кластера, с учетом доступности серверов с вероятностью r имеем

$$P = (1-P_1)\left[\sum_{i=a}^{n} C_n^i p_0^i (1-p_0)^{n-i}\right] + P_1 \left[\sum_{i=1}^{n} C_n^i p_0^i (1-p_0)^{n-i}\right] \sum_{j=b}^{m} C_m^j p_2^j (1-p_2)^{m-j},$$

где $p_2 = rp_0$. Для систем критического применения, не допускающих наличие узлов, отказ которых может вызвать отказ системы, в качестве базовых средств вычислений используются резервированные вычислительные комплексы [12]. Простейшая структура дублированного вычислительного комплекса (ДВК), скомплектованная из двух связанных через адаптер сопряжения (АС) полукомплексов, включающих процессоры (П) и модули памяти (М), представлена на рис. 2, а. Модель надежности ДВК, допускающего возможность совместной работы процессора и модуля памяти разных полукомплексов, сводится к хорошо изученной в теории надежности модели мостиковой схемы [10, 11], приведенной на рис. 2, б.

Надежность (коэффициент готовности) ДВК, в соответствии с моделью по рис. 2, б, вычисляется как

$$p_0 = p_a(1-(1-p_p)^2)(1-(1-p_M)^2) + (1-p_a)(1-(1-p_p p_M)^2),$$

где при заданных интенсивностях отказов $\lambda_p, \lambda_M, \lambda_a$ и восстановлений μ_p, μ_M, μ_a процессора, памяти и адаптера сопряжения соответственно имеем $p_p = \mu_p / (\lambda_p + \mu_p)$, $p_M = \mu_M / (\lambda_M + \mu_M)$, $p_a = \mu_a / (\lambda_a + \mu_a)$.

В случае невозможности совместной работы процессоров и модулей памяти разных полукомплексов надежность ДВК вычислим как

$$p_0 = (1-(1-p_p p_M)^2).$$

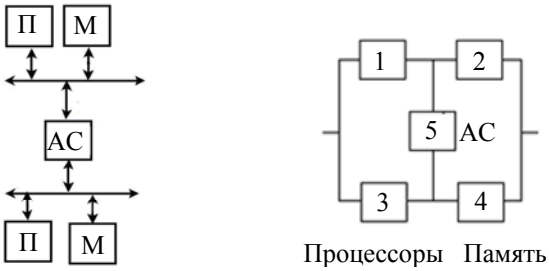

Рис. 2. Структура (а) и модель надежности (б) ДВК

Для ДВК с ограниченным восстановлением (одновременный ремонт нескольких узлов невозможен) коэффициент готовности определяется как сумма вероятностей работоспособных состояний, для нахождения которых процесс отказов и восстановлений представляется марковским процессом, при этом составляется граф переходов и уравнения Чепмена–Колмогорова, в результате решения которых и определяются искомые вероятности. При оценке вероятностей работоспособных состояний и коэффициента готовности ДВК по рис. 2, а, могут использоваться результаты, полученные в [13].

Оптимизация структуры

При оптимизации структуры рассматриваемой вычислительной системы ищется число серверов n в локальных кластерах, число серверов m в общедоступном кластере и кратность резервирования N коммутационных узлов, обеспечивающие максимум надежности системы, $P = \max_{m,n,N,g} P(m,n,N,g,\lambda)$, при ограничении стоимости s ее реализации $(Mc_0 n + c_1 N + c_2 m) \le s$, и условия стационарности функционирования узлов (отсутствия перегрузки узлов).

Поиск максимума P может основаться на переборе, реализуемом с использованием средств системы компьютерной математики Matchcad-15.

Целью оптимизации структуры может быть минимизация среднего времени пребывания запросов в системе [14] при ограничении средств s на ее построение, $T = \min_{m,n,N,g} T(m,n,N,g,\lambda)$, при этом среднее время пребывания запросов в системе вычисляется [9] как

$$T = g\left(\frac{v_0}{1-\frac{g\lambda v_0}{n}}\right)+(1-g)\left(\frac{2v_1}{1-\frac{\left((1-g)+\beta\right)2\lambda v_1}{N}}+\frac{v_0}{1-\frac{\left((1-g)+\beta\right)\lambda v_0}{m}}\right), \quad (2)$$

где $(1-g)$ – средняя доля запросов, перераспределяемых через сеть от локального кластера в общедоступный. При поиске оптимального g необходимо учитывать условие стационарного режима функционирования узлов (условие отсутствия перегрузки узлов) [9]:

$$\left(\frac{g\lambda v_0}{n}<1\right)\wedge\left(\frac{\left((1-g)+\beta\right)2\lambda v_1}{N}<1\right)\wedge\left(\frac{\left((1-g)+\beta\right)\lambda v_0}{m}<1\right). \quad (3)$$

При необходимости оптимизация может быть проведена по мультипликативному критерию $r(m,n,N,g,\lambda)=\max_{m,n,N,g}\left(P(m,n,N)/T(m,n,N,g,\lambda)\right)$.

Оптимизация процесса перераспределения запросов

При заданной структуре системы (сформированной при рассмотренной выше структурной оптимизации) проведем оптимизацию процесса распределения запросов с учетом возможности отказов и отключений исправных узлов общедоступного кластера с вероятностью $(1-r)$. Оптимизация проводится при заданной средней интенсивности потока запросов λ, поступающего в локальный кластер и при необходимости перераспределяемого через сеть в общедоступный кластер.

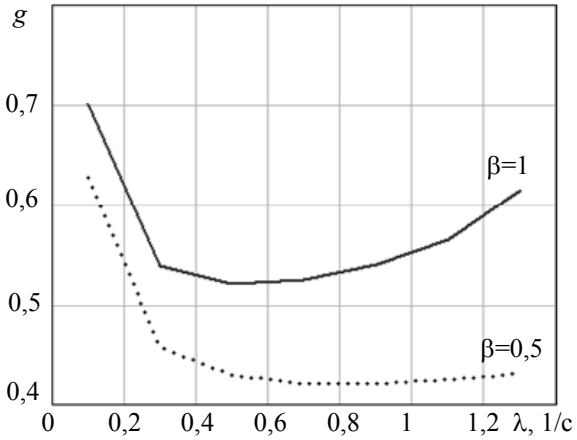

Рис. 3. Оптимальная доля запросов, перераспределяемых через сеть

В результате оптимизации процесса распределения потока запросов, поступающего в локальный кластер, ищется их доля, перераспределяемая через сеть в общедоступный кластер, при которой минимизируется среднее время пребывания запросов T. $T = \min_{g} T(m,n,N,g,\lambda)$, где при модернизации (2) и (3) имеем

$$T = g\left(\frac{v_0}{1-\frac{g\lambda v_0}{n}}\right)+(1-g)\left(\frac{2v_1}{1-\frac{\left((1-g)+\beta\right)2\lambda v_1}{N_c}}+\frac{v_0}{1-\frac{\left((1-g)+\beta\right)\lambda v_0}{m_c}}\right),$$

$$\left(\frac{g\lambda v_0}{n}<1\right)\wedge\left(\frac{\left((1-g)+\beta\right)2\lambda v_1}{N_c}<1\right)\wedge\left(\frac{\left((1-g)+\beta\right)\lambda v_0}{m_c}<1\right)$$

при математических ожиданиях числа коммутационных узлов N_c и доступных исправных серверов общедоступного кластера, вычисляемых как

$$N_c = \sum_{i=1}^{N} iC_N^i p_1^i (1-p_1)^{N-i}\ ,\ m_c = \sum_{j=1}^{m} jC_m^j p_2^j (1-p_2)^{m-j}\ ,$$

$$\left(\frac{g\lambda v_0}{n} < 1\right) \wedge \left(\frac{\left((1-g)+\beta\right)2\lambda v_1}{N_c} < 1\right) \wedge \left(\frac{\left((1-g)+\beta\right)\lambda v_0}{m_c} < 1\right).$$

Для примера проведем оптимизацию процесса распределения запросов при $n = 8$ шт., $N = 5$ шт., $m = 23$ шт.; $v_0 = 10$ с, $v_1 = 1$ с, $r = 0{,}8$; $\lambda_0 = \lambda_2 = 10^{-4}$ 1/ч, $\lambda_1 = 0{,}5 \cdot 10^{-4}$ 1/ч; $\mu_0 = \mu_1 = \mu_2 = 1$ 1/ч. Результаты поиска оптимальной доли $(1 - g)$, распределяемых через сеть в общедоступный кластер запросов, в зависимости от интенсивности входного потока запросов λ 1/с представлены на рис. 3 при $\beta = 0{,}5$ и $\beta = 1$. Рост доли неперераспределяемых запросов g при незначительной интенсивности λ потока запросов объясняется влиянием дополнительных задержек при передаче запросов через сеть, а при значительной интенсивности λ – перегрузкой общедоступного кластера.

Заключение

Поставлены и решены задачи оптимизации структуры вычислительной системы и процесса перераспределения через сеть потока запросов от локальных кластеров в общедоступный кластер с учетом возможностей отказов, восстановлений и отключений серверов общедоступного кластера. Перераспределение запросов реализуется с целью минимизации среднего времени пребывания запросов при адаптации системы к отказам узлов и изменениям потока запросов.

Предложены модели надежности и массового обслуживания вычислительных систем динамического перераспределения запросов (нагрузки) между кластерами, которые могут быть использованы при оценке надежности и выборе рациональных вариантов организации перераспределения запросов в системах с объединением вычислительных ресурсов в локальные и общедоступные кластеры, связанные через сеть.

Работа выполнена на кафедре вычислительной техники НИУ ИТМО в рамках НИР «Разработка методов и средств системотехнического проектирования информационных и управляющих вычислительных систем распределенной архитектуры».

Литература

1. Таненбаум Э., Ван Стеен М. Распределенные системы. Принципы и парадигмы. – СПб: Питер. – 2003. – 877 с.
2. Clark T. The New Data Center. New technologies are radically reshaping the data center. – Brocade Bookshelf. San Jose, 2010. – 156 p.
3. Кармановский Н.С., Гатчин Ю.А., Терентьев А.О., Федоров Д.Ю., Беккер М.Я. Информационная безопасность при облачных вычислениях: проблемы и перспективы // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 1 (71). – С. 97–102.
4. Богатырев В.А. К повышению надежности вычислительных систем на основе динамического распределения функций // Изв. вузов. Приборостроение. – 1981. – № 8. – С. 62–65.
5. Богатырев В.А. Распределение заданий в многомашинных вычислительных системах // Изв. вузов. Приборостроение. – 1986. – № 5. – С. 43–47.
6. Богатырев В.А. Надежность функционально-распределенных резервированных структур с иерархической конфигурацией узлов // Изв. вузов. Приборостроение. – 2000. – № 4. – С. 67–70.
7. Богатырев В.А. Надежность вычислительных систем с функциональной реконфигурацией на основе перераспределения задач // Информационные технологии. – 2001. – № 7. – С. 22–27.
8. Богатырев В.А., Богатырев С.В. Объединение резервированных серверов в кластеры высоконадежной компьютерной системы // Информационные технологии. – 2009. – № 6. – С. 41–47.
9. Bogatyrev V.A., Golubev I.Y., Bogatyrev S.V. Optimization and the Process of Task Distribution between Computer System Clusters // Automatic Control and Computer Sciences. – 2012. – V. 46. – № 3. – P. 103–111.
10. Гуров С.В., Половко А.М. Основы теории надежности. – СПб: БХВ-Петербург, 2006. – 704 с.
11. Черкесов Г.Н. Надежность аппаратно-программных комплексов. – СПб: Питер, 2005. – 479 с.
12. Bogatyrev V.A. Exchange of Duplicated Computing Complexes in Fault tolerant Systems // Automatic Control and Computer Sciences. – 2011. – V. 46. – № 5. – P. 268–276.
13. Богатырев В.А., Башкова С.А., Беззубов В.Ф., Голубев И.Ю., Котельникова Е.Ю., Полякова А.В. Надежность дублированных вычислительных комплексов // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 6. – С. 74–78.
14. Алиев Т.И. Основы моделирования дискретных систем. – СПб: СПбГУ ИТМО, 2009. – 363 с.

Богатырев Владимир Анатольевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, Vladimir.bogatyrev@gmail.com

Богатырев Анатолий Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, gangleon@gmail.com Vladimir.bogatyrev@gmail.com

Голубев Иван Юрьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, www.golubev@mail.ru

Богатырев Станислав Владимирович – ООО «Айти Хаус», главный инженер; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, realloc@gmail.com Vladimir.bogatyrev@gmail.com