

## АНАЛИЗ РАБОТЫ УЗЛОВ РАСПРЕДЕЛЕННЫХ СИСТЕМ ОБРАБОТКИ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ GERT-СЕТЕЙ

Цепкова М.И.<sup>1,2</sup>, Ступина А.А.<sup>1,2</sup>, Корпачева Л.Н.<sup>1</sup>, Федорова А.В.<sup>1</sup>, Джиоева Н.Н.<sup>1</sup>

<sup>1</sup> ФГАОУ ВПО «Сибирский федеральный университет» (Красноярск, Россия (660025, Красноярск, пер. Вузовский, 3)), e-mail: kafedraIT@mail.ru

<sup>2</sup> ФГБОУ ВПО «Сибирский государственный аэрокосмический университет имени академика М.Ф. Решетнёва» (Красноярск, Россия (660014, Красноярск, пр. им. газ. «Красноярский рабочий, 31), e-mail: saa55@rambler.ru

---

Для большого класса задач важным является возможность использования надежной системы обработки информации с приемлемой производительностью и возможностью одновременной обработки нескольких наборов данных. Одним из видов таких систем являются распределенные гетерогенные вычислительные отказоустойчивые системы для высокопроизводительных вычислений. В статье приведен аналитический обзор распределенных гетерогенных вычислительных отказоустойчивых систем для высокопроизводительных вычислений, представлен математический аппарат модифицированных GERT-сетей для анализа временных характеристик работы узлов распределенной гетерогенной системы обработки, позволяющий проводить расчет стохастической GERT-сети с использованием произвольного числа дополнительных вещественных и стохастических параметров узла стохастической GERT-сети, предложены прямой и обратный алгоритмы их расчета. Произведено аналитическое сравнение их производительности.

Ключевые слова: GERT-сеть, распределенная гетерогенная система обработки, высокопроизводительные вычисления.

## ANALYSIS OF WORK UNITS OF THE DISTRIBUTED INFORMATION PROCESSING SYSTEM USING GERT-NETWORKS

Tsepkova M.I.<sup>1,2</sup>, Stupina A.A.<sup>1,2</sup>, Korpacheva L.N.<sup>1</sup>, Fedorova A.V.<sup>1</sup>, Dzhioeva N.N.<sup>1</sup>

<sup>1</sup> Siberian federal university (Krasnoyarsk, Russia (660025, Krasnoyarsk, Vuzovskii, 3)), e-mail: kafedraIT@mail.ru

<sup>2</sup> Siberian State Aerospace University named after M.F. Reshetnev, Krasnoyarsk, Russia (660014, Krasnoyarsk, Krasnoyarski Rabochiy, 31), e-mail: saa55@rambler.ru

---

It is important to the use the reliable information processing system with the acceptable performance and the ability to process multiple data sets for a large class of problems. One type of such systems is fail-safe distributed heterogeneous computing systems for high performance computing. The article considers an analytical overview of the distributed heterogeneous computing fault-tolerant systems for high performance computing, presented the mathematical apparatus modified GERT-network for the analysis of the temporal characteristics of the work units of the distributed heterogeneous processing system, that allows to estimate stochastic GERT-network using an arbitrary number of additional real and stochastic parameters of the node stochastic GERT-network offered direct and inverse algorithms for their estimation. The analytical comparison of their performance is given.

Keywords: GERT-network, distributed heterogeneous processing system, high performance computing.

В настоящее время все острее встают вопросы обеспечения отказоустойчивости систем обработки информации. Этому есть несколько причин. Во-первых, системы с каждым годом усложняются, включают в себя все большее количество компонентов, а, следовательно, вероятность того, что в одном из компонентов системы произойдет сбой, увеличивается. Это может привести к отказу всей системы, если она не спроектирована с учетом такой возможности.

Для большого класса задач не требуется высокая производительность обработки одного набора данных. Гораздо важнее возможность использования надежной системы

обработки информации с приемлемой производительностью и возможностью одновременной обработки нескольких наборов данных.

На сегодняшний день системы обработки информации все чаще используются для решения управленческих, исследовательских и производственных задач. Одним из видов таких систем являются гетерогенные вычислительные системы высокой надежности для высокопроизводительных вычислений (системы обработки высокой пропускной способности). Наиболее известные технологии: Legion, Condor, Apples PST, Netsolve, Punch, XTRemweb и т.д. используют простые схемы распределения, когда центральный компьютер, отвечающий за распределение, решает, какие задачи должны быть выполнены на каком ресурсе, используя функции стоимости, задаваемые системными параметрами. Они не рассматривают цену использования каждого ресурса, а это означает, что значимость выполнения всех приложений в любое время одинакова, что в реальности далеко не так. Значимость должна возрастать с приближением срока выполнения прикладной задачи.

Гетерогенность состава вычислительных узлов и непредсказуемые изменения вычислительной среды во время решения задачи приводят к проблеме рационального использования вычислительной мощности, сосредоточенной в сети.

Анализ существующих на сегодняшний день технологий и программных средств обработки информации показывает, что процесс разработки приложений с использованием сети в качестве вычислительного ресурса является сложным, т.к. содержит множество этапов, начиная от разработки параллельного алгоритма и заканчивая мониторингом ресурсов и распределения нагрузки.

Проблема распределения нагрузки в параллельных вычислениях является одной из самых важных. И именно от решения этой проблемы в основном зависит эффективность параллельного решения задачи, т.е. тот выигрыш во времени, который можно получить по сравнению с последовательным решением.

Некоторые из рассматриваемых средств реализуют методы распределения нагрузки, что снижает трудоемкость разработки приложений при использовании этих средств. Но данные методы не обеспечивают рационального использования ресурсов сети, т.к. не исключают простоя компьютеров. Так, например, в ADM (ApplicationDataMovement) точность распределения зависит от некоторой функции, которую должен написать разработчик. Метакомпьютинг Condor требует описания ресурсов от разработчика и распределение нагрузки производит на основании этого описания. При этом не учитывается реальная загрузка компьютеров, которая может меняться во время вычисления. Метакомпьютинг Piranha просто выбирает один из свободных компьютеров случайным образом, не учитывая тем самым реальных возможностей компьютеров и требуемых

ресурсов для задачи. SunGridEngine и Netsolve распределение нагрузки производят на основании данных мониторинга сети, где собирают информацию о загрузке вычислительных узлов в текущий момент времени и соответственно с этим отправляют задачу на наименее загруженный компьютер. Подобный подход обеспечивает более точное распределение нагрузки, не требующее от разработчика каких-либо данных или действий, но при этом возникают дополнительные накладные расходы на осуществление мониторинга.

Рассмотренные вышеуказанные метакомпьютинги не обеспечивают оптимального использования вычислительных ресурсов сети с учетом минимизации времени решения задачи и накладных расходов [2].

Большинство исследовательских задач направлены на создание методов обработки информации и управления системой обработки информации, стремящихся приблизить практическую продолжительность работы (производительность, коэффициент ускорения) к теоретически возможной для данной системы обработки информации. Для этого создаются и совершенствуются системы сбора и обработки информации о работе вычислительной системы и методы оценки предельной практической и теоретической производительности системы обработки информации для конкретных задач.

При анализе эффективности работы, производительности и управляемости системы обработки информации возникает задача оценки временных характеристик работы системы или ее узлов, а также оценки времени выполнения задания в различных режимах запуска при различных условиях. Для оптимизации загрузки (составления расписания) системы обработки информации, как правило, используются методы теории оптимизации. Для моделирования поведения системы и оценки временных характеристик системы применяются сети Петри и их модификации. При исследовании системы с учетом возможной ненадежности узлов, как правило, применяются Марковские цепи или стохастические сети.

При анализе временных характеристик работы узлов распределенной гетерогенной системы обработки, использующей в качестве узлов персональные компьютеры пользователей во время их «простоя», данные методы становятся неприменимы. Такие системы не статичны во времени. Для их узлов важными характеристиками являются «доступность» и «отказоустойчивость» [1].

Поэтому для решения поставленной задачи используются модифицированные GERT-сети, позволяющие использовать все шесть типов узлов; проводить оценку временных характеристик стохастической GERT-сети с использованием произвольного числа дополнительных вещественных и стохастических параметров узла стохастической GERT-

сети; использовать в качестве условных вероятностей выполнения исходящей дуги узла произвольные функции, вычислимые в момент активации узла.

Пусть направленный граф  $G$  состоит из непустого множества  $E$  направленных ребер (дуг) и непустого множества  $V$  вершин (узлов). Пусть каждое ребро однозначно определяется по его начальному и конечному узлам (ребро  $(i, j)$  определено начальным узлом  $i$  и конечным  $j$ ).

Каждый узел сети активируется с некоторой вероятностью. Для стохастической GERT-сети весом дуги  $(i, j)$  является вектор  $[p_{ij}, F_{ij}]$ , где  $p_{ij}$  – условная вероятность выполнения дуги  $(i, j)$  при условии активации узла  $i$  (вероятность выполнения дуги (работы)  $(i, j)$ ), а  $F_{ij}$  – условная функция распределения времени выполнения дуги  $(i, j)$ , при условии, что  $(i, j)$  выполняется.

Каждый узел сети имеет входную и выходную функции активации, также влияющие на параметры активируемого узла.

Виды входных функций.

AND-функция – узел активируется, если выполнены все дуги, входящие в него.

IOR-функция – узел активируется, если выполнена любая дуга, входящая в него.

EOR-функция – узел активируется, если выполнена любая дуга, входящая в него, при условии, что в данный момент времени может выполняться только одна дуга, входящая в данный узел.

Виды выходных функций.

Детерминированная функция – все дуги, выходящие из узла, выполняются, если узел активирован.

Стохастическая функция – ровно одна дуга, выходящая из узла, выполняется с заданной вероятностью, если узел активирован.

Комбинируя все входные и выходные функции, получаем шесть различных типов узлов.

Активация узла означает, что система перешла в некоторое состояние и определяет множество возможных дальнейших действий. Одно или несколько действий начинают свое выполнение сразу после активации узла, являющегося их началом. Активация узла происходит, если его входная функция выполнена. После выполнения выходной функции активированного узла (начала выполнения соответствующей дуги) он становится неактивным.

GERT-сеть – это сеть с источниками  $R$  и стоками  $S$  вида «работа на дуге», в которой каждый узел принадлежит одному из шести типов узлов, для каждой дуги  $(i, j)$  определен вес вида  $[p_{ij}, F_{ij}]$  с вышеуказанным значением и задано начальное распределение источников сети.

Введем обозначения:

$$P(i) = \{j \in v \mid (i, j) \in E\};$$

$$S(i) = \{j \in v \mid (j, i) \in E\};$$

$$R(i) = \{\text{множество узлов достижимых из узла } i \text{ (включая } i)\};$$

$$\bar{R}(i) = \{\text{множество узлов, из которых узел } i \text{ достижим (включая } i)\}.$$

Обозначим  $G(i)$ , где  $i \in R$  – подсеть сети  $G$ , построенная на множестве вершин  $R(i)$ .

Обозначим  $\Psi$  множество последовательностей активации сети.

Множество последовательностей активации  $\Psi$  называется допустимым, если для любых  $W_1, W_2 \in \Psi$ ,  $W_1 \neq W_2$ , пути  $W_1$  и  $W_2$  не пересекаются, исключая случай, если они используют один и тот же детерминированный начальный узел и могут использовать один и тот же конечный узел.

Пусть  $D_{ij}^n$  – время выполнения в  $n$ -ый раз дуги  $(i, j)$ .

MG-сетью называется GERT-сеть  $G(V, E)$  если:

1.  $G$  имеет единственный источник и, по крайней мере, один сток.
2. Сеть  $G$  удовлетворяет ограничениям:
  - а. В течение каждого выполнения проекта для каждого стока активируется не более одного источника, из которого данный сток достижим.
  - б. Для каждого узла  $i$  MG-сети, если узел  $i$  активирован, то параметры всех выходящих из него дуг вычислимы.

Для расчета параметров узлов с AND- и IOR-входами требуется не только рассчитать параметры всех входящих дуг, но и знать вероятность активации узла, «породившего» процесс одновременного выполнения событий.

Пусть узел  $i \in V$  GERT-сети  $G(V, E)$  имеет более одного предка ( $|P(i)| > 1$ ), тогда для любых  $j, k \in P(i)$  введем функцию  $Pr(i, j, k)$  – множество узлов, являющихся ближайшими общими предками для узла  $i$  и путей, заканчивающихся дугами  $(j, i)$ ,  $(k, i)$ :

$$Pr(i, j, k) = \{l \mid l \in \bar{R}(j) \cap \bar{R}(k), S(l) \cap \bar{R}(j) \cap \bar{R}(k) = \emptyset\}.$$

Аналогичным образом для узла  $i \in V$  GERT-сети  $G(V, E)$ , имеющего более одного потомка ( $|S(i)| > 1$ ), для любых  $j, k \in S(i)$  введем функцию  $Sc(i, j, k)$  – множество узлов, являющихся ближайшими общими потомками для узла  $i$  и путей, начинающихся дугами  $(i, j)$ ,  $(i, k)$ :

$$Sc(i, j, k) = \{l \in V \mid l \in R(j) \cap R(k), P(l) \cap R(j) \cap R(k) = \emptyset\}.$$

в. Для каждого узла  $k$  произвольной циклической структуры  $C$  существует путь из  $k$  к узлу вне  $C$ , такой, что  $p_{ij} > 0$  для каждой дуги  $(i, j)$  данного пути.

г. Для всякого узла  $i$  GERT-сети  $G$ , имеющего IOR- или AND-вход, для любых  $j, k \in P(i)$ ,  $Pr(i, j, k) = \{1\}$ , причем 1 – единственный узел и 1 имеет детерминированный выход.

д. Для всякого узла  $i$  GERT-сети  $G$ , имеющего детерминированный вход, для любых  $j, k \in S(i)$ ,  $Sc(i, j, k) = \{1\}$ , причем 1 – единственный узел и 1 имеет AND- или IOR- вход.

е. Реализация сети является допустимой, если в процессе выполнения каждый из активированных узлов сети активируется не более чем 1 раз; реализация сети является допустимой, если в процессе выполнения каждый из активированных узлов сети активируется с вероятностью, большей  $P_{max} > 0$ ; если в узел  $i$  циклической структуры  $C$  входит более одной дуги и хотя бы одна дуга не принадлежит циклу  $C$ , то узел  $i$  имеет EOR-вход; если из узла  $i$  циклической структуры  $C$  выходит более одной дуги и хотя бы одна дуга не принадлежит циклу  $C$ , то узел  $i$  имеет стохастический выход; если узел  $i$  с IOR- или AND-входом принадлежит циклу, то узел  $j$  с детерминированным выходом, являющийся стохастическим началом узла  $i$ , принадлежит данному циклу. Данное гарантирует, что для любой сети количество реализаций конечно, следовательно, сеть вычислима с использованием ЭВМ.

3. Задано множество параметров, для каждого узла сети (по крайней мере, вероятность активации).

4. Для каждой дуги указаны функции преобразования параметров узлов сети.

5. Источник активируется в момент времени 0.

Под стохастическими параметрами узлов мы всегда подразумеваем функции распределение вещественных случайных величин.

Результатом расчета сети будет множество реализаций  $R_r$ , сгруппированных по стокам  $r_k$ . Для стока в каждой реализации будет рассчитана вероятность активации узла (вероятность возникновения реализации), функция распределения времени выполнения данной реализации и все дополнительные параметры. Пусть  $r$  – элемент множества реализаций  $R_r$  со стоком  $r_k$ .

Вероятность активации стока  $r_k$ :  $P_{rk} = \sum_r p_r$ .

Вероятность активации стока  $r_k$  ко времени  $t$  при условии активации стока  $r_k$  (функция распределения времени выполнения стока  $r_k$ ):  $F_{rk}(t) = p_{rk}(t)/P_{rk} = \sum_r p_r * F_r(t) / P_{rk}$ .

В качестве дополнительных переменных в MG-сети могут выступать вещественные или стохастические переменные. Для каждой вещественной переменной определено ее начальное значение в источнике сети, и каждая дуга содержит функцию преобразования значения переменной. Для каждой стохастической переменной определено ее начальное распределение в источнике сети, и каждая дуга содержит условную функцию распределения

некоторой случайной величины и указание на одну из операций: « + », « - », « = » (присвоить) [3].

Результатом работы прямого и обратного алгоритмов расчета МG-сети является множество реализаций, удовлетворяющих ограничению  $e$ .

Общее описание обратного алгоритма расчета МG-сети:

- обход графа МG-сети ведется от выбранного стока к источнику;
- обход ведется с помощью рекурсивного алгоритма;
- если узел имеет EOR-вход, то для каждой входящей дуги запускается рекурсивный вызов процедуры обхода сети;
- если узел  $i$  имеет IOR- или AND-вход, то запускается рекурсивный вызов процедуры расчета узла  $j$ , являющегося детерминированным источником узла  $i$ , затем рассчитываются все пути от  $j$  к  $i$  и, используя найденные пути, строятся все реализации, завершаемые узлом  $i$ ;
- расчет вещественных и стохастических параметров производится при «выходе» алгоритма из рекурсии.

Общее описание прямого алгоритма расчета МG-сети:

- обход графа МG-сети ведется от источника к стокам;
- обход ведется с помощью рекурсивного алгоритма;
- если узел имеет стохастический выход, то для каждой выходящей дуги запускается рекурсивный вызов процедуры обхода сети;
- если узел  $i$  имеет детерминированный выход, то рассчитываются все пути от  $i$  к  $j$ , и, используя найденные пути, строятся все реализации, завершаемые узлом  $j$ , где  $j$  – детерминированный сток узла  $i$ ;
- расчет вещественных и стохастических параметров производится при «углублении» алгоритма рекурсии.

Функции распределения заданы множеством узлов на равномерной сетке. Для аппроксимации производной функции распределения использована формула первого порядка, поскольку она дает строго неотрицательные значения функций плотности распределения. Для численного интегрирования интеграла свертки в зависимости от выбранного режима расчета используется либо квадратурная формула трапеции, либо быстрое преобразование Фурье [4].

По результатам аналитического сравнения прямой алгоритм работает не медленнее обратного, однако требует выполнения ограничения  $e$ .

Таким образом, произвольная математическая модель МG-сети с шестью типами узлов может быть рассчитана на ЭВМ.

## Список литературы

1. Гергель, А.В., Виноградов, Р.В. Оценка сложности коммуникационных операций в кластерных вычислительных системах // Высокопроизводительные параллельные вычисления на кластерных системах. Материалы второго Международного научно-практического семинара / под ред. проф. Р.Г. Стронгина. – Нижний Новгород: Изд-во Нижегородского госун-та. – 2002. – С. 73-77.
2. Мирошник, М.А. Синтез распределенных вычислительных сред на базе компьютерных сетей // Системы обработки информации. – 2013. – № 7 (114). – С. 86-89.
3. Письман, Д.М. Анализ временных параметров сетевых моделей на базе модифицированной ГЕРТ-сети. // Проблемы машиностроения и автоматизации. – 2006. – № 1. – С. 18-26.
4. Письман, Д.М. Сравнение производительности прямого и обратного алгоритмов расчета модифицированной ГЕРТ-сети. // Фундаментальные исследования. – 2006. – № 2. – С. 45-47.
5. Ступина, А.А. Планирование задач в распределенных гетерогенных информационных системах // Современные проблемы науки и образования. – 2013. – № 5.

### Рецензенты:

Антамошкин А.Н., д.т.н., профессор, профессор кафедры экономики и информационных технологий менеджмента института управления бизнес-процессами и экономики ФГАОУ ВПО «Сибирский федеральный университет» (Министерство образования и науки РФ), г. Красноярск;

Попов А.М., д.ф.-м.н., профессор, директор института информатики и телекоммуникаций ФГБОУ ВПО «Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева» (Министерство образования и науки РФ), г. Красноярск.

## **Краткие сведения об авторах**

**Цепкова Мария Игоревна**, 1990 г.р, ассистент Сибирского Федерального Университета  
Область научных интересов: теория систем и системный анализ, математическое моделирование.

Служебный адрес: 660025, г. Красноярск, пер. Вузовский, 3, тел.: (3912) 06-37-61

Домашний адрес: 660001, Красноярск, ул. Менжинского, д. 6, кв. 176.

e-mail: marie1206@yandex.ru

**Ступина Алена Александровна**, 1973 г.р., д-р техн. наук, профессор Сибирского Федерального Университета. Область научных интересов: теория систем и системный анализ, моделирование бизнес-процессов.

Служебный адрес: 660025, г. Красноярск, пер. Вузовский, 3, тел.: (3912) 06-37-61.

Домашний адрес: 660032, Красноярск, ул. Дубенского, д. 4ж, кв. 56.

e-mail: h677hm@gmail.com

**Корпачева Лариса Николаевна**, 1960 г.р., канд. техн. наук, доцент Сибирского Федерального Университета. Область научных интересов: моделирование бизнес-процессов, экономико-математическое моделирование, теория систем и системный анализ.

Служебный адрес: 660025, г. Красноярск, пер. Вузовский, 3, тел.: (3912) 06-37-61.

Домашний адрес: 660097, г. Красноярск, ул. Дубровинского, д. 58, кв. 69.

e-mail: korp\_0777@mail.ru

**Федорова Александра Витальевна**, 1966 г.р., доцент Сибирского Федерального Университета, кандидат геол.-минерал. наук. Область информационные технологии управления, анализ и усовершенствование бизнес-процессов

Служебный адрес: 660025, г. Красноярск, пер. Вузовский, 3, тел.: (3912) 06-37-61

Домашний адрес: 660059, Красноярск, ул. Вавилова, д. 47б, кв. 311.

e-mail: alvitfedorova@gmail.com

**Джиоева Наталья Николаевна**, 1965 г.р., доцент Сибирского Федерального Университета, кандидат технических наук. Область научных интересов: математическое моделирование, методология реинжиниринга, имитационное моделирование.

Служебный адрес: 660025, г. Красноярск, пер. Вузовский, 3, тел.: (3912) 06-37-61

Домашний адрес: 660036, Красноярск, Академгородок, д. 26, кв. 3.

e-mail: DzhioevaNatalia@gmail.com