

УДК 519.168+519.687

МОДЕЛЬ ЗАДАЧИ ФОРМИРОВАНИЯ СООБЩЕСТВ УСТРОЙСТВ ИНФОРМАЦИОННО-УПРАВЛЯЮЩИХ СИСТЕМ В СРЕДАХ ТУМАННЫХ ВЫЧИСЛЕНИЙ

Э.В. Мельник

Федеральный исследовательский центр Южный научный центр Российской академии наук
344006, г. Ростов-на-Дону, пр. Чехова, 41
E-mail: Evm17@mail.ru

А.Б. Клименко

Научно-исследовательский институт многопроцессорных вычислительных систем им. А.В. Каляева
(НИИ МВС) Южного федерального университета
347922, ЮФО, Ростовская область, г. Таганрог, ул. Чехова, 2
E-mail: anna_klimenko@mail.ru

Д.Я. Иванов

Научно-исследовательский институт многопроцессорных вычислительных систем им. А.В. Каляева
(НИИ МВС) Южного федерального университета
347922, ЮФО, Ростовская область, г. Таганрог, ул. Чехова, 2
E-mail: donat.ivanov@gmail.com

Ключевые слова: информационно-управляющая система, распределение задач, расписание, оптимизация, туманные вычисления, облачные вычисления.

Аннотация: данная статья посвящена вопросу формализации задачи размещения вычислительной нагрузки высоконадежной информационно-управляющей системы в случае функционирования последней в среде туманных вычислений. Актуальность работы обусловлена тем, что в условиях повсеместного внедрения концепций «Интернет вещей» и «промышленный интернет вещей», все большее распространение получают платформы туманных вычислений, реализующие принципы перемещения нагрузки в рамках последней. Постановка задачи отличается от аналогов, сформулированных для гетерогенных вычислительных сред, поскольку учитывает такие важные характеристики туманного слоя как расстояние между устройствами, а также параметры, описывающие подграф задачи, необходимый к размещению.

1. Введение

Географически распределенные информационно-управляющие системы (ИУС) в условиях интенсивного расширения и интеграции, что соответствует концепции Industry 4.0, в настоящее время используют не только собственную сетевую инфраструктуру и вычислительную среду, но и имеющиеся в наличии каналы связи общего назначения, а также вычислительные ресурсы датацентров [1, 2]. Это приводит к зависимости качества функционирования ИУС от методов и средств, используемых платформами, в

рамках которых запускаются приложения: например, при отсутствии специальной информации об особенностях конфигурации, система Kubernetes разместит приложение в соответствии с собственными приоритетами [3]. Такой подход не всегда оптимален для ИУС, к которым предъявляются высокие требования в плане надежности и времени отклика [4, 5], поскольку размещение частных реализаций вычислительных подзадач могут существенно влиять как на скорость решения комплексной задачи управления, так и на надежность вычислительных узлов. Это же актуально и для систем, функционирование которых предполагается в средах туманных вычислений по причине гетерогенности инфраструктуры и ее географической распределенности [6-8], в отличие от кластера датацентра. Концепция туманных вычислений предполагает наличие возможности переноса вычислительной нагрузки из «облачного слоя», реализованного на основе инфраструктуры датацентра, в слой «туманный», состоящий из коммуникационного оборудования различных типов. Однако, как уже было сказано выше, платформа, реализующая концепцию «туманных» вычислений, должна быть осведомлена о том, что, если речь идет о функционировании ИУС с повышенными требованиями к надежности, необходимо учитывать предпочтительные варианты размещения вычислительных подзадач с целью повышения показателей надежности. Возможна предварительная подготовка потенциальных вариантов перераспределения вычислительной нагрузки, но это не всегда целесообразно, во-первых, по причине количества таких вариантов, во-вторых, по причине динамики структуры туманного слоя. По этой причине наиболее приемлемым способом размещения вычислительных задач является динамическое формирование сообществ устройств, на которых могут быть запущены информационно взаимосвязанные задачи.

Целью данной статьи является разработка модели задачи формирования сообществ устройств для решения комплексов информационно-связанных подзадач в рамках задачи управления распределенной ИУС, функционирующей на основе инфраструктуры туманных вычислений. В последующих разделах будет изложена модель задачи размещения комплекса информационно-связанных подзадач с акцентом на специфику среды туманных вычислений, а также предложены методы динамического формирования сообществ устройств для их решения.

2. Размещение комплекса информационно связанных подзадач в среде туманных вычислений

К настоящему времени разработан широкий круг моделей задач о распределении вычислительных задач по системе исполнителей, в том числе, гетерогенной и в том числе, с учетом формирования множества устройств. Примеры таких постановок даны в работах [9-11]. Также для распределения задач по однородным вычислительным ресурсам зачастую используется модель задачи двумерной упаковки [12]. Однако, наличие среды туманных вычислений привносит необходимость модификации существующих моделей с целью отражения ее специфики и специфики ставящейся задачи, что в рамках модели, может быть сведено к следующему:

- поскольку речь идет о задаче управления, состоящей из информационно связанного комплекса подзадач, имеет значение структура подграфа, определяющего переносимую нагрузку, включая количественные характеристики, такие как объемы пересылаемых между подзадачами данных, вычислительная сложность подзадач и их связность с другими задачами в графе, которые не переносятся;

- концепция туманных вычислений предполагает выход за пределы облачного слоя, что продуцирует новый параметр – расстояние между вычислительными устройствами. Кроме того, поскольку туманный слой распределен географически, вычислительные устройства могут быть разнесены относительно друг друга таким образом, что между ними будут присутствовать транзиторные устройства, через которые будет происходить передача данных.
- поскольку целью данной работы является повышение надежности устройств, используемых ИУС, одним из параметров модели задачи становится показатель надежности ВБР (вероятность безотказной работы), на который, в том числе, влияет объем передаваемой и принимаемой узлом информации.

Исходный комплекс подзадач описывается графом G информационно связанных подзадач с вычислительными сложностями x_i и ребрами, взвешенными объемами передаваемой между задачами информации w_{ik} . Пусть подграф G' должен быть переразмещен в пределах сегмента сети P' (будем считать его ограниченным), в то время как задачи подграфа G'' продолжают быть закреплены за узлами подмножества P'' , как показано на рис. 1.

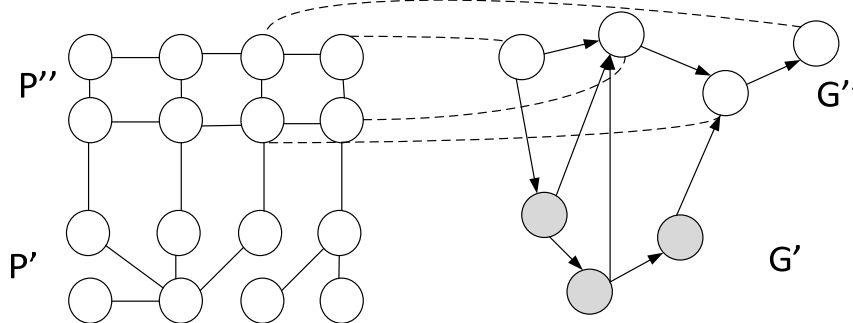


Рис. 1. Задача переразмещения G' по P' .

Переразмещение предполагает наличие критерия оценки качества размещения, т.е. целевой функции, скалярной или векторной.

Представим модель задачи формирования сообществ устройств высоконадежных информационно-управляющих систем в средах туманных вычислений в формализованном виде.

Граф G представляет собой описание комплекса информационно связанных подзадач, которые должны быть решены: $G = \{ \langle i, x_i, w_i \rangle \}$, где i – идентификатор подзадачи, x_i – вычислительная сложность задачи, w_i – объем данных, передаваемых i -й задачей в коммуникационную среду.

При этом изначально задачи G привязаны к узлам множества P , где P также описывается графовой структурой $P = \{ \langle j, p_j \rangle \}$, где j – идентификатор узла, p_j – производительность узла, $list$ – матрица, определяющая пропускную способность каналов связи между смежными узлами сети.

Предположим, что вычислительная нагрузка, соответствующая подграфу G' должна быть перемещена, в то время как подзадачи, соответствующие подграфу G'' , что составляет разность между G и G' , остается для выполнения на прежних местах. При этом между подграфами G' и G'' остаются информационные связи между подзадачами, которые, относительно перемещаемого графа G'' могут быть описаны в виде множеств кортежей следующего вида:

$Flow_in = \{ \langle id_{out}, id_{in}, w_{out_in} \rangle \}$ – это множество кортежей, каждый из которых описывает объемы передаваемых данных между внешним по отношению к G' узлом id_{out} и принадлежащим G' узлом id_{in} .

$Flow_out = \{ \langle id_{in}, id_{out}, w_{in_out} \rangle \}$ – это множество кортежей, каждый из которых описывает объемы передаваемых данных между внутренним узлом id_{in} и принадлежащим G'' узлом id_{out} .

Сформулируем задачу следующим образом: имея G'' прикрепленным к P'' , а также $Flow_in, Flow_out$, необходимо разместить G' по P' таким образом, чтобы время решения G укладывалось в отведенный период T , и при этом достигался максимум целевой функции (за которую в рамках данной работы принимается вероятность безотказной работы (ВБР) $_j, j=1..M$, где M – число узлов в P' , является примером ЦФ). Решением описываемой задачи будет являться закрепление подзадач G' за P' , в виде матрицы A :

$$A = \begin{pmatrix} \langle t_0^{ij}, u_{ij} \rangle & & \\ & \dots & \\ & & \langle t_0^{NM}, u_{NM} \rangle \end{pmatrix},$$

где t_0^{ij} – момент времени начала выполнения задачи i узлом j , u_{ij} – доля производительности p_j , выделяемая узлом j для решения задачи i .

Отметим, что в рамках данной модели допускается выполнение одним узлом нескольких задач, что некоторым образом расширяет классическую модель составления расписаний [10].

Для того, чтобы произвести размещение подграфа G' по ограниченному подграфу P' с учетом особенностей систем туманных вычислений и ВБР устройств, необходимо, помимо $Flow_in$ и $Flow_out$, ввести следующие дополнительные параметры:

- $L_p(A)$ – загруженность, порождаемая решением распределенными на узел вычислительными задачами;
- $L_{dist}(A, Flow_in, Flow_out)$ – загруженность, порождаемая необходимостью устройства принимать данные для выполнения закрепленных на нем задач, и передавать данные в вычислительную среду в рамках G' и непосредственно в G'' ;
- $L_{tr}(A, Flow_in, Flow_out)$ – загруженность, порождаемая транзитной передачей данных через узел;
- D_{lk} – данный параметр содержит список ребер P , определяющий маршрут между узлами l и k ;
- $ListD_{lk}$ – матрица, определяющая пропускные способности каналов связи между узлами l и k соответственно.

Пусть в качестве ЦФ выбираются ВБР некоторых устройств, которые могут быть приоритетны. Обозначая $ВБР_j$ через F_j , получим равенство следующего вида:

$$F_j = e^{-\lambda_j t},$$

где λ_j – интенсивность отказов j -го устройства на этапе стационарного функционирования, t – время функционирования устройства.

Поскольку $\lambda = \lambda_0 \cdot 2^{\Delta T/10}$, а $\Delta T = kL$, где L – загруженность устройства, k – коэффициент пропорциональности и зависит от типа устройства, получим зависимость ВБР устройства от его загруженности:

$$F_j = e^{\lambda_{j0} \cdot 2^{kL/10}}.$$

Загруженность устройства, в свою очередь, зависит от способа распределения задач, определяемого матрицей A и включает следующие компоненты: $L_p(A)$ – загруженность, порождаемая решением распределенными на узел вычислительными задачами; $L_{dist}(A, Flow_in, Flow_out)$ – загруженность, порождаемая необходимостью устройства принимать данные для выполнения закрепленных на нем задач, и передавать данные в вычислительную среду в рамках G' и непосредственно в G'' ; $L_{tr}(A)$ – загруженность, порождаемая транзитной передачей данных через узел;

Таким образом, суммарная загруженность устройства j , продуцируемая выполняющимися на нем задачами, будет иметь вид:

$$L_j = L_{pj}(A) + L_{distj}(A, Flow_in, Flow_out) + L_{trj}(A).$$

Подставляя полученное выражение загруженности устройства в выражение ВБР _{j} , получим ЦФ для одного устройства. В случае необходимости оптимизации ВБР нескольких устройств, формируется векторная ЦФ.

Основным ограничением в рамках данной работы принимается ограничение на время выполнения всего комплекса задач, т.е.:

Для $G = G' \cup G''$,

$$\forall i \in G: \frac{x_i}{p_j u_{ij}} + t_{dist}(i) < T,$$

где $t_{dist}(i)$ – максимальное время доставки данных от задачи i к задачам-приемникам информации.

Поскольку в модели задачи учитываются такие параметры, как маршруты передачи информации между задачами, привязанными к узлам, параметр $t_{dist}(i)$ вычисляется при помощи функционала, реализующего расчет времени передачи информации задаче-приемнику, расположенной на произвольном узле:

$$t_{dist}(i) = \xi(A, G, P),$$

т.е. времена доставок информации вычисляются на основе полной информации о размещении подзадач относительно друг друга и с использованием параметров D_{lk} и $ListD_{lk}$.

3. Заключение

Несмотря на широкий круг публикаций в предметной области моделирования распределения вычислительной нагрузки и составления расписаний, как правило, модели не учитывают таких особенностей распределенных систем, которые порождаются спецификой сред туманных вычислений:

- географическую распределенность вычислительных устройств, на которых могут быть решены вычислительные задачи;
- дополнительную вычислительную нагрузку на узлы, которая появляется в результате необходимости передавать данные по сети через узел и также влияет на надежность характеристики узлов.

В предлагаемой модели задачи формирования сообществ устройств высоконадежных информационно-управляющих систем в средах туманных вычислений учтены перечисленные особенности функционирования ИУС в средах туманных вычислений.

Выбранный критерий оптимизации – БВР устройств – позволяет сформировать требования к инфраструктуре среды туманных вычислений для размещения вычислительной нагрузки в узлах туманного слоя с учетом требований к показателям надежности ИУС.

Поскольку вычисление ЦФ и ограничений весьма затруднительно с помощью классических методов исследования операций и может быть произведено алгоритмически, решение поставленной задачи целесообразно производить методами итерационного поиска, например, алгоритмами на основе метода имитации отжига или генетическими алгоритмами.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (17-08-01605, 18-29-22046) и ГЗ ЮНЦ РАН № гр. проекта АААА-А19-119011190173-6.

Список литературы

1. <https://www.i-scoop.eu/industry-4-0/scada-supervisory-control-data-acquisition/>
2. http://aig.si/13/clanki/vabljena_predavanja/U5_Dragisic.pdf
3. <https://kubernetes.io/>
4. Каляев И.А., Мельник Э.В. Децентрализованные системы компьютерного управления. Ростов н/Д: Издательство ЮНЦ РАН, 2011. 196 с.
5. Melnik E., Korovin I., Klimenko A. Improving dependability of reconfigurable robotic control system, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017. doi: 10.1007/978-3-319-66471-2_16.
6. Wang Y., Uehara T., Sasaki R. Fog computing: issues and challenges in security and forensics // Proceedings of the International Computer Software and Applications Conference. 2015. P. 53-59. doi: 10.1109/COMPSAC.2015.173.
7. Hu P. et al. Survey on fog computing: architecture, key technologies, applications and open issues // Journal of Network and Computer Applications. 2017. P. 27-42. doi: 10.1016/j.jnca.2017.09.002.
8. Alrawaiis A. et al. Fog Computing for the Internet of Things: Security and Privacy Issues // IEEE Internet Computing. 2017. Vol. 21, No. 2. P. 34-42. doi: 10.1109/MIC.2017.37.
9. Барский А.Б. Параллельные процессы в вычислительных системах. Планирование и организация. М.: Радио и связь, 1990.
10. Pinedo M. L. Scheduling: Theory, algorithms, and systems, fifth edition, Scheduling: Theory, Algorithms, and Systems / Fifth Edition. 2016. doi: 10.1007/978-3-319-26580-3.
11. Гончар Д.Р., Фуругян М.Г. Эффективные алгоритмы планирования вычислений в многопроцессорных системах реального времени // УБС. 2014. №49. URL: <https://cyberleninka.ru/article/n/effektivnye-algoritmy-planirovaniya-vychisleniy-v-mnogoprotsessornyh-sistemah-realnogo-vremeni> (дата обращения: 19.11.2018).
12. Martello S., Toth P. Bin-Packing problem // Knapsack problems: Algorithms and computer implementations. 1990.