

Мультиагентные технологии управления ресурсами в распределенных вычислительных средах

А.В. Прохоров, Е.М. Пахнина

Национальный аэрокосмический университет им. Н.Е. Жуковского, Харьков, Украина

avprohorov@yahoo.com, elena.pakhnina@khai.edu

Аннотация. *С развитием кластерных и облачных вычислений возникают все большие сложности в координации и распределении ресурсов. Предлагаемый в работе подход, основанный на мультиагентных технологиях, позволяет автоматизировать полный цикл управления ресурсами в реальном времени, включая согласование интересов, скоординированное взаимодействие, динамическое планирование и адаптивное перепланирование запросов/ресурсов. Описаны механизмы информационного взаимодействия агентов, связанные с распространением волны переговоров при конфликте интересов в виде локализации и динамического перераспределения нагрузки во внутриузловых и межузловых коммуникациях.*

Ключевые слова

Управление ресурсами, мультиагентные системы, компетенция, динамическое перераспределение нагрузки

1 Введение

Сегодня все чаще при организации вычислений применяется разделение работ, данных и процессорных мощностей, а также иные режимы взаимодействия, предусматривающие использование распределенных ресурсов [1]. Управление ресурсами в сложных динамических средах представляет собой процесс определения требований, сопоставления ресурсов и приложений, распределения ресурсов, планирования, балансировки нагрузки и мониторинга, реализуемый в middleware систем кластерных и облачных вычислений. При этом растет сложность и динамика принятия решений по управлению ресурсами: неопределенность: непредсказуемая динамика спроса/предложения/доступности ресурсов; событийность: постоянно происходят события, которые требуют внесения корректив в планы вычислений; ситуативность: отсутствие типовых решений, а необходимость их принятия по ситуации; многофакторность: наличие множества разных критериев, политик, предпочтений и ограничений, что приводит к необходимости балансировки между ними; взаимозависимость: принятие одного решения вызывает изменение других; индивидуальность: потребители услуг требуют все более индивидуального подхода; конфликты: все больше участников с противоречивыми интересами, требующих скоординированного, многоуровневого планирования ресурсов; трудоемкость: большое число и разный характер факторов для того, чтобы просчитать последствия того или иного варианта при сохранении высокой оперативности.

Все это формирует современную постановку задачи динамического управления ресурсами в распределенных системах (рис. 1). Имеется множество ресурсов, причем комплексный состав ресурсов уникален для каждого конкретного предоставляемого сервиса, поэтому необходимо моделировать их состав и взаимодействие. При этом следует учитывать, что ресурсы и запросы на них кооперируются, образуя группы посредством применения согласованных стратегий, временная логика которых зависит от динамически меняющихся условий. В реальном времени поступают запросы, задачи и любые другие события (задержки, отказы и т.д.), которые необходимо планировать, учитывая текущие планы, индивидуальные предпочтения и ограничения заказов и ресурсов. При этом изменения должны вноситься в планы ресурсов без останова и перезапуска, путем адаптивного изменения расписания «на лету» с использованием как свободных временных «окон», так и трансфером ранее распределенных запросов. Должен быть реализован полный цикл управления: реакция на события, динамическое планирование, согласование и пересмотр планов «на лету», мониторинг и

контроль исполнения планов, а в случае расхождения плана и факта требуется автоматическое перепланирование. Следует также учитывать, что ресурсы могут работать во взаимодействии, влиянии одних на другие в системе, то есть, создавая дополнительные, синергетические свойства. И, наконец, система ресурсов должна быть сбалансирована, то есть ресурсы должны соответствовать общей стратегии по количеству, качеству обслуживания, производительности, так как общая их результативность будет определяться самым слабым местом.

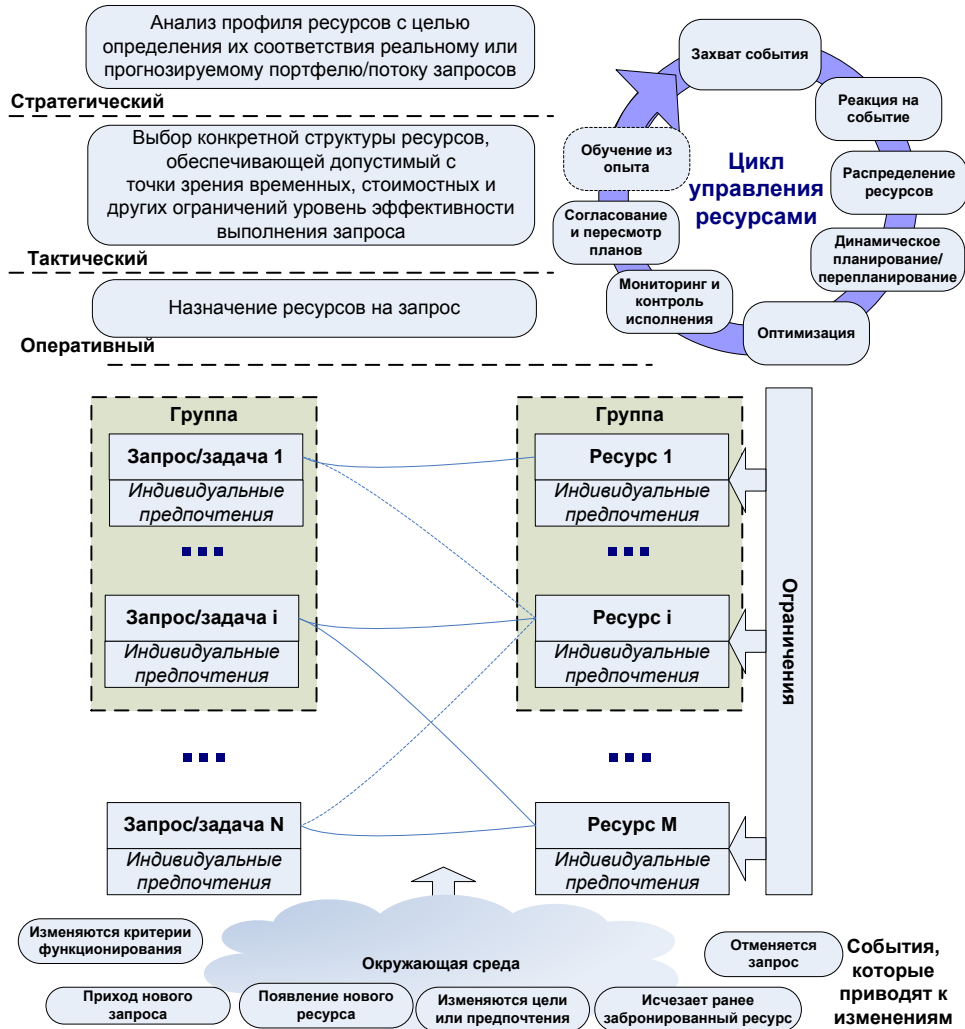


Рис. 1. Задача динамического управления ресурсами в распределенных системах

В современных параллельных архитектурах и кластерных системах используются [2,3]:

- *планирование вычислений на уровне приложений.* Роль посредников между пользователями и вычислительными узлами выполняют брокеры ресурсов. Однако, использование независимыми пользователями различных критериев для оптимизации планов выполнения своих заданий (в условиях возможной конкуренции с другими заданиями) может ухудшать такие интегральные характеристики распределенной среды, как время выполнения пакета заданий и загрузка ресурсов;
- *планирование на уровне потоков заданий.* Метапланирование строится на иерархической (древовидной) структуре управления заданиями в виртуальной организации, состоящей из метапланировщика, подконтрольных ему менеджеров заданий, которые, в свою очередь, взаимодействуют с локальными менеджерами управления ресурсами или системами пакетной обработки заданий. Здесь время выполнения отдельных приложений может увеличиваться, поскольку при планировании не удается учесть важных особенностей структуры заданий и пользовательских предпочтений, касающихся ресурсов.

Анализ показал, что в настоящее время в подходах к управлению ресурсами не учитывается динамика изменения загрузки узлов; не учитывается конкуренция независимых пользователей, а также глобальных и локальных потоков заданий собственников ресурсов; не в полной мере поддерживаются оптимизационные механизмы планирования заданий.

В то же время повышения эффективности управления ресурсами в распределенных вычислительных системах можно достичь путем внедрения в механизмы работы их middleware методов и средств интеллектуальных информационных технологий. В качестве базового элемента такой интеллектуализации grid или cloud вычислений может рассматриваться семантический информационный сервис, построенный на основе онтологии грид-ресурсов и представляющий собой мультиагентную систему [4]. Целью данного исследования является рассмотрение вопросов повышения эффективности использования ресурсов и загрузки в сложных распределенных средах на основе самоорганизации и адаптации в мультиагентных системах в условиях неопределенности и динамического окружения.

2 Мультиагентная модель управления ресурсами

2.1 Основные положения

Наиболее эффективный анализ и прогнозирование динамических характеристик при исследовании реализуемости портфеля/потока запросов/заказов при ограниченных ресурсах обеспечивают средства, основанные на методах имитационного моделирования. При этом для обеспечения гибких механизмов динамического поведения, автономности и адаптации отдельных компонентов имитационной модели целесообразным является использование агентно-ориентированного подхода. Мультиагентные системы (МАС) с учетом предоставляемых ими преимуществ — автономности, децентрализованности, индивидуального поведения с возможностью обучения и адаптации для составляющих их агентов — автоматизируют полный цикл управления ресурсами в реальном времени, включая согласование интересов, скоординированное взаимодействие, динамическое планирование и адаптивное перепланирование запросов/ресурсов.

В рамках данного исследования предлагается мультиагентная имитационная модель управления ресурсами в сложных динамических средах, с возможностями экономической самоорганизации и коллективной адаптации элементов модели с собственными интересами и процессами принятия решений, которая позволяет осуществлять децентрализованное планирование с гибким управлением ресурсами и нагрузкой в условиях неопределенности и динамического окружения.

Множество ресурсов и запросов/заказов моделируются в виде агентов, каждый узел в этом случае представляет собой агентную платформу, под управлением которой функционируют агенты заказов и ресурсов. Отличительной особенностью предложенной модели является введение в составе характеристик ресурсов множества компетенций, которыми располагает тот или иной ресурс, а в составе заказов набора компетенций, которые необходимы для их выполнения. В условиях интеграции кластерных/облачных технологий и сервисно-ориентированной архитектуры SOA можно провести аналогию между компетенцией и сервисом.

Рассмотрим основные составляющие модели.

Имеется множество ресурсов $R = \{R_1, R_2, \dots, R_M\}$. Каждый ресурс R_m ($m = \overline{1..M}$) характеризуется:

- ценой s_m , которая выражается в условных денежных единицах и назначается самим ресурсом в зависимости от ситуации на рынке ресурсов;
- заданной себестоимостью в расчете на единицу времени c_m , которая не изменяется, причем, обычно $s_m > c_m$;
- подмножеством $H_m^R = (h_{m1}^R, h_{m2}^R, \dots, h_{mk}^R)$, которое представляет собой набор компетенций, которыми располагает ресурс, h_{mi}^R - характеризует доступные (свободные) мощности по каждой компетенции;
- подмножеством $E_m^R = (e_{m1}^R, e_{m2}^R, \dots, e_{mk}^R)$, которое представляет собой эффективность (способность) ресурса при выполнении компетенции h_{mi}^R ;
- прибыль ресурса $SV(R_m) = \sum_{i=1}^n (s_m(O_i) - c_m(O_i))$ рассматривается как сумма прибылей от выполнения каждого заказа, где $s_m(O_i)$ - стоимость ресурса для заказа O_i , а $c_m(O_i)$ - себестоимость выполнения этого заказа;
- вектором индивидуальных предпочтений $PV(R_m)$ - это предпочтения ресурса по времени выполнения заказа (краткосрочные и долгосрочные заказы), объему заказа и т.д.

Рассматривается множество узлов $WN = \{WN_1, WN_2, \dots, WN_L\}$. Каждый узел WN_m ($m = \overline{1..L}$) характеризуется

- вектором свойств $PV(WN_m)$ - это характеристики вычислительных узлов;

- подмножеством ресурсов, таким, что каждый ресурс $R_m = \{UWN_j | WN_j \cap OWN\} \cap WN$

Выделяется множество заказов $O = \{O_1, O_2, \dots, O_N\}$. Каждый заказ O_n ($n = \overline{1..N}$) характеризуется:

- комплексом работ $Y = \{y_{nj}\}$, где $j = \overline{1..p_n}$ где p_n - количество работ n-го заказа. Соответственно план работ по заказу будет представлять собой множество работ $y_{nj} = (t_{nj}^f, t_{nj}^e, s_{nj}, d_{nj})$, где t_{nj}^f, t_{nj}^e - сроки начала и окончания работы y_{nj} , s_{nj} - объём финансовых ресурсов, выделенный для выполнения работы, d_{ij} - стандартная продолжительность работы;

- подмножеством $H_n^O = (h_{n1}^O, h_{n2}^O, \dots, h_{nk}^O)$ - набор компетенций, которые необходимы для выполнения заказа, h_{ni}^O

- характеризует требуемые производственные мощности (или уровень) по каждой компетенции;

- покупательной способностью p_n , выражается денежными единицами;

- объемом средств (резервом) r_n , которые заказ может потратить на получение ресурса в данный момент времени, причем $r_n < p_n$;

- показателем соответствия целям системы SC_n ;

- прибыль заказа $SV(O_n) = \sum_{i=1}^n (r_n(R_i) - s_i)$ определяется как сумма денежных средств от использования ресурсов для выполнения заказа, т.е. разница между объемом средств $r_n(R_i)$, имеющихся на покупку i-го ресурса и его текущей ценой s_i , принятой заказом.

- вектором индивидуальных предпочтений $PV(O_m)$ - это предпочтения заказа относительно желательного и предельно допустимого срока завершения, объема, штрафа за невыполнение заказа и т.д.

В составе мультиагентной модели управления ресурсами каждый агент-менеджер *OrderManager* представляет собой конкретный заказ/задачу и характеризуется своими целями, приоритетом, сроками, стоимостью. Целью для каждого агента *OrderManager* является поиск контрактов с ресурсами, которые могут выполнить необходимые работы с минимальными затратами. Для этого *OrderManager* будет рассчитывать локальное оптимальное расписание с учетом затрат на ресурсы, как результат имитации рыночных взаимодействий между заказами и ресурсами.

Ресурсы являются элементом с индивидуальным поведением, поэтому моделируются агентом *ResourceManager*. Его цель состоит в том, чтобы увеличить уровень загрузки ресурсов и доходов.

2.2 Особенности взаимодействия агентов модели управления ресурсами

Наибольшую сложность в теоретических исследованиях и практических реализациях современных МАС представляют вопросы, связанные с процессами взаимодействия агентов при коллективном решении задач реальной практической сложности и значимости, поскольку каждый агент, решающий конкретную подзадачу, имеет лишь частичное представление об общей задаче и должен постоянно взаимодействовать с другими агентами. В связи с этим в настоящее время наиболее актуальной в теории и технологии МАС является проблема создания моделей взаимодействия интеллектуальных агентов в МАС.

Агенты обмениваются друг с другом информацией посредством переговоров. Взаимодействие агентов в нашем исследовании организуем таким образом, чтобы максимизировать глобальную привлекательность портфеля заказов. Глобальную привлекательность портфеля/потока заказов/запросов предлагается оценивать путем усреднения полезности (учитывает финансовый результат SV и эффект, как соответствие целям системы SC), полученной в определенном временном интервале.

Эффективное решение в системе появляется в результате применения аукционного механизма. Каждый агент *OrderManager* создает свой собственный локальный план, принимая во внимание свои собственные цели и знания. Полученные таким образом локальные планы в общем случае являются несовместимыми (например, несколько заказов пытаются использовать одновременно один и тот же ресурс). Более того, локальные планы могут быть глобально неэффективны («прибыльные» заказы могут быть

отклонены, наиболее важные задержаны и т.д.). Эти проблемы, которые являются результатом автономности каждого агента, будем решать посредством введения рыночного механизма, который бы гарантировал максимальную совместимость локальных планов и глобальную эффективность.

Задача одновременного выделения ресурсов для независимых параллельных заданий решается на основе локальных расписаний, представляющих собой списки *слотов* – отрезков времени, которые могут быть выделены заданиям при соответствующих требованиях к уровню и типам вычислительных ресурсов локальных узлов.

В каждом цикле планирования запуска портфеля запросов/заказов требуется: отбор подходящих (по ресурсу, времени, цене) слотов; выбор комбинации слотов, являющейся оптимальной с точки зрения прохождения всего портфеля заданий в текущем цикле планирования.

Для реализации аукционного подхода в мультизаказном планировании в данной работе используем метод множителей Лагранжа. Каждый агент *OrderManager* играет роль претендента, участвующего в аукционах, запрашивая у агентов *ResourceManager* набор временных интервалов – слотов, необходимых для выполнения текущей задачи, ожидающей выполнения из комплекса работ по заказу. Таким образом, он будет пытаться отыскать множество временных интервалов через ресурсный пул, в соответствии с минимально возможной локальной стоимостью. В состав этой стоимости включены две составляющих: суммарная стоимость выделенных ресурсных временных интервалов и штрафная стоимость задержки сроков выполнения работы. Агент, играющий роль аукциониста, определяет цену, оплачиваемую за интервалы времени ресурсов с целью снижения ресурсных конфликтов и максимизации их дохода. Чтобы достичь этой цели он использует для корректирования цены в каждом круге торгов алгоритм субградиентной оптимизации.

2.3 Динамическое распределение нагрузки между узлами

При межузловом взаимодействии программных агентов-задач возникает ряд проблем, связанных с возрастанием нагрузки на сеть, уменьшением интенсивности информационного обмена между агентами, зависящей от скорости и пропускной способности каналов связи сети, увеличением времени доставки сообщений и времени поиска нужного агента для взаимодействия.

Решение – объединение интенсивно взаимодействующих агентов в *коалиции* [5] и преобразование межузлового взаимодействия агентов во взаимодействие на одном общем узле. Агенты-задания группируются и направляются в различные домены вычислительных узлов. Однако при этом допускается трансфер: задания, для которых, например, не удается найти подходящих ресурсов в текущем цикле планирования, могут направляться в другие домены узлов.

Второй важный аспект – объединение под управлением одного менеджера тех вычислительных ресурсов, которые являются схожими по архитектуре, составу, политике администрирования. Решение также является объединение сходных по интересам агентов в *коалиции* или группы.

На основе подхода, предложенного в [6] определим основными агентными механизмами, обеспечивающими повышение эффективности работы системы управления ресурсами следующие:

- *локализация агентных взаимодействий* в пределах узлов путем формирования групп активно коммуницирующих агентов
- *динамического перераспределения нагрузки* за счет реализации механизмов группового трансфера агентов.

На каждом узле функционируют следующие агенты (рис. 2):

- *системный монитор (SystemMonitor)*, периодически проверяющий нагрузку на текущем узле компьютерной сети;
- *менеджер распределения агентов (DistributedAgent)*, ответственный за динамическое распределение агентов и анализ моделей поведения и характера взаимодействия агентов;
- *менеджер трансфера агентов (TransferAgent)*, перемещающий агенты и группы агентов на другие узлы и управляющий трансфером агентов между узлами сети.

Менеджер распределения агентов *DistributedAgent* оценивает интенсивность межузловых и внутриузловых коммуникаций агентов системы.

Когда максимальное значение коэффициента коммуникационной зависимости агента больше предопределенного порога, *DistributedAgent* текущего узла включает рассматриваемого агента в группу агентов, расположенную на удаленном узле системы. Перед перемещением выделенной группы агентов с текущего узла на удаленный узел проверяется текущее состояние памяти, загруженность центрального процессора и число агентов, размещенных на данном узле, с помощью системного монитора. Если удаленный узел имеет достаточное количество свободных системных ресурсов для новых агентов, то инициируется перемещение выбранной группы агентов на принимающий узел. Механизм локализации межагентных взаимодействий

предполагает анализ динамических изменений в моделях взаимодействия агентов, но этот механизм может вызвать перегрузку некоторых из узлов системы из-за большого числа агентов, перемещаемых на эти узлы.

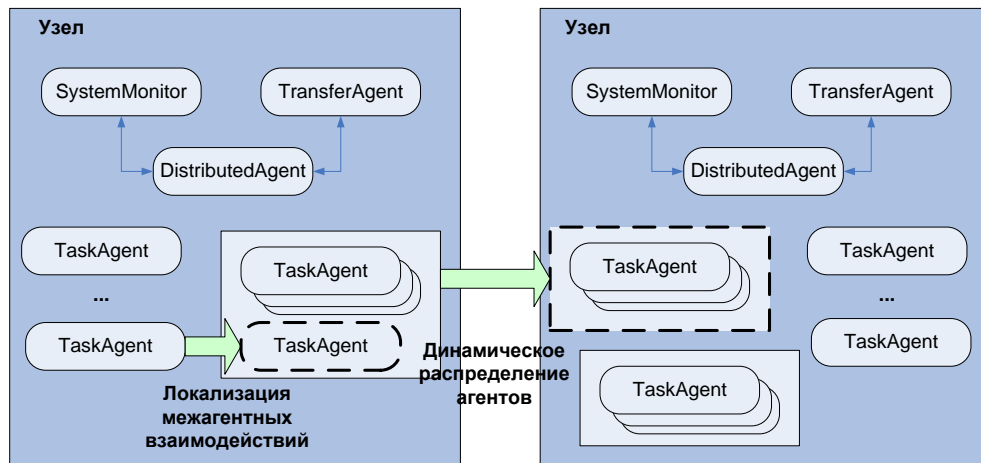


Рис. 2. Динамическое распределение нагрузки между узлами

Когда узел перегружен, системный монитор обнаруживает это состояние и активирует процедуру перераспределения агентов, основанную на перемещении не отдельных агентов между узлами сети, а целых групп агентов.

В продолжение рассмотрения взаимодействия агентов при коллективном решении задач были разработаны модели формирования коалиции агентов на основе анализа компетенций, требуемых для успешного выполнения возлагаемых на нее задач. Предложен метод решения с использованием линейной оптимизационной модели и варианты ее использования: формирование группы с минимальным числом агентов-ресурсов; формирование группы с минимальной стоимостью.

Одним из вариантов модели является ее использование при формировании устойчивой коалиции. Под устойчивостью коалиции понимаем ее способность сохранять полный перечень компетенций при удалении из нее одного или нескольких участников. Решение данной задачи осуществляется с помощью алгоритма уменьшения требований. Так если не найдены кандидаты по полному набору требуемых компетенций, то рассматривается минимально необходимый набор компетенций в зависимости от важности (уменьшается размерность пространства для поиска кандидатов в гиперкубе), но в любом случае рассчитывается расстояние по полному набору требуемых компетенций.

Для того, чтобы проводить оценку эффективности информационного взаимодействия и управлять стратегиями поведения агентов целесообразно проводить учет следующих свойств или критериев: *осторожность, самостоятельность, активность, управление, ресурс, кооперативность, конфликтность*. Таким образом, стратегия поведения агента будет определяться в зависимости от степеней проявления указанных свойств, которые будут оцениваться по следующим шкалам:

- «осторожности – риска» («осторожный» агент торгуется и в результате находит наилучшее предложение, а «рискованный» – немедленно резервирует первое же предложение, если оно удовлетворяет его требованиям);
- «активности – пассивности» (определяется частотой выдвижения предложений, выполнения роли инициатора задач);
- «самостоятельность-зависимость» (определяется отношением полученных и подготовленных управляющих воздействий);
- «централизации–децентрализации» управления (показывает тяготение агента к директивному или коллективному стилю управления);
- «централизации – децентрализации» ресурсов (характеризует готовность агента предоставлять ресурсы в общее использование или, наоборот, занимать чужие ресурсы);
- «кооперативности – конкурентности» (характеризует агента с точки зрения формирования объединений, в случае нехватки ресурсов);
- «согласия – конфликтности» (если эта оценка высока, то агент расположен часто расторгать соглашения, если же низка, то агент склонен к длительному сотрудничеству).

Таким образом, системообразующими информационными технологиями, которые взаимодополняют друг друга могут быть использованы для решения задач динамического распределения ресурсов являются мультиагентные технологии, онтологии, сервис-ориентированная архитектура и семантик веб. Их интеграция

ведет к созданию нового класса интеллектуальных распределенных систем, ориентированных на кооперирование посредством взаимодействия сервисов, предоставляемых и потребляемых интеллектуальными агентами.

На основе предложенных методов и моделей разработана компьютерная система управления ресурсами. Практическая реализация выполнена в виде прототипа знаниеориентированной системы на основе агентного подхода, на базе платформы JADE, в составе которой функционируют интеллектуальные агенты, осуществляющие принятие решений и взаимодействие с помощью онтологической базы знаний и механизма логического вывода. Основу системы составляет семантический информационный сервис, построенный на основе онтологии грид-ресурсов [4]. Все агенты системы используют разработанную базовую онтологию ресурсов грид, которая содержит специфические для грид фундаментальные концепты и отношения. Т.е. целью создания базовой онтологии является обеспечение гибкости и адаптивности семантического информационного сервиса и возможности его применения в различных грид-архитектурах и middleware. Можно выделить два основных класса задач управления ресурсами, решение которых должно опираться на онтологию: поиск, сопоставление профилей компетенций с помощью мер схожести (например, для сопоставления профиля компетенций ресурса с профилем требований задачи); выявление необходимости поиска конкретного ресурса (для этого необходимо выявить разрыв – отсутствие компетенций - путем сравнения профиля требований задачи с текущим профилем компетенций ресурсов). При расчете меры семантической близости при этом используется гибридный метод, который содержит оценку близости экземпляров, состоящую из трех частей – таксономической, реляционной и атрибутивной. Использование онтологии таким образом решает задачу формирования единого информационного пространства управления ресурсами. В настоящее время проводятся экспериментальные исследования при различных стратегиях планирования по экономическим принципам самоорганизации агентов: минимизация времени выполнения задания при ограничении на бюджет; максимизация доходов собственников ресурсов при ограничении на суммарное время использования временных слотов; минимизация суммарной стоимости выполнения заданий при ограничении на суммарное время использования слотов. Предварительные результаты показывают эффективность использования мультиагентных технологий при управлении ресурсами в сложных динамических средах.

3 Заключение

Предложена мультиагентная имитационная модель управления ресурсами в кластерных и облачных вычислительных средах, с возможностями экономической самоорганизации и коллективной адаптации элементов модели с собственными интересами и процессами принятия решений, которая позволяет осуществлять децентрализованное планирование с гибким управлением ресурсами и нагрузкой в условиях неопределенности и динамического окружения. Описаны механизмы информационного взаимодействия агентов, связанные с распространением волны переговоров при конфликте интересов в виде локализации и динамического перераспределения нагрузки во внутриузловых и межузловых коммуникациях агентов. Рассматриваются модели динамического формирования коалиций агентов за счет учета компетенций, требуемых для успешного выполнения возлагаемых на них задач пользователей грид-системы.

Список литературы

- [1] Foster I.T., Kesselman C., Nick J.M., Tuecke S. Grid Services for Distributed System Integration // IEEE Computer. 35(6). – 2002. – pp. 37-46.
- [2] Волк М.А., Филимончук Т.В., Гридель Р.Н. Методы распределения ресурсов для Grid-систем // Сборник научных работ Харьковского университета Воздушных Сил. – 2009. Вып. 1(19). – с. 100 – 104.
- [3] Коваленко В.Н., Коваленко Е.И., Корягин Д.А., Семячкин Д.А. Управление параллельными заданиями в гриде с неотчуждаемыми ресурсами. // Препринт №63. – М.: ИПИ РАН. – 2007. – с. 1 –28.
- [4] Prokhorov A.V., Pakhnina E.M. Agent-based Semantic Grid Service for Resource Management // Proceedings of International Conference "Parallel and Distributed Computing Systems" (PDCS'2013). – pp. 249-256.
- [5] Шереметов Л.Б. Модель нечетких коалиционных игр в задачах конфигурирования открытых сетей поставок // Известия РАН. Теория и системы управления. 2009. № 5. с. 94–108.
- [6] Маслобоев А.В. Модели и алгоритмы взаимодействия программных агентов в виртуальной бизнес-среде развития инноваций // Вестник МГТУ, том 12, №2. 2009. – с. 224-234.