

ИМИТАЦИОННАЯ МОДЕЛЬ ЦЕНТРА ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ В ВИДЕ СИСТЕМЫ С ГРУППОВЫМ ПОСТУПЛЕНИЕМ ЗАЯВОК

Закирова Р.И., Михалкин Д.В.

Российский университет дружбы народов, renatazakirova@gmail.com, dvmixalkin@mail.ru

Построена имитационная модель центра облачных вычислений с использованием нескольких систем массового обслуживания типа M M c/r.

Ключевые слова: облачные вычисления, время отклика, вероятность блокировки.

Введение

Облачные вычисления – новая парадигма предоставления вычислительной инфраструктуры конечному потребителю, цель которой снизить затраты на управление и техническое обслуживание аппаратных и программных ресурсов [1].

Одно из ключевых требований к среде облачных вычислений - обеспечение гарантированного качества обслуживания (Quality of service, QoS), которое в терминах соглашения об уровне предоставления услуг (Service level agreement, SLA) можно описать такими характеристиками как пропускная способность, время отклика и доступность. При организации центров облачных вычислений (ЦОВ) перед поставщиками стоит задача предоставления услуг с высоким QoS и с оптимальным распределением ресурсов.

В работе построена имитационная модель центра облачных вычислений (ЦОВ) согласно [2,3], которая позволяет оценить основные характеристики функционирования ЦОВ - вероятность блокировки запроса пользователя на предоставление услуги облачных вычислений и время отклика ЦОВ на запрос пользователя, а также некоторые другие характеристики функционирования ЦОВ, напр., среднюю длину очереди подзапросов к каждому поставщику услуг.

Описание модели

Центр облачных вычислений, объединяющий серверы K поставщиков услуг, предоставляет пользователям M типов услуг. При предоставлении услуг каждого типа задействуются ресурсы нескольких поставщиков, при этом наборы поставщиков для разных услуг могут различаться. Поступающий в ЦОВ запрос пользователя попадает в узел-планировщик, где разбивается на несколько подзапросов в соответствии с определенным для предоставления услуги данного типа набором поставщиков, подзапросы направляются на обслуживание к соответствующим поставщикам услуг, обслуживаются на серверах поставщиков, после чего все обслуженные подзапросы собираются в узле-сборщике. Запрос пользователя на предоставление услуги считается обслуженным, когда обслужены все составляющие его подзапросы. Таким образом, время отклика ЦОВ на запрос пользователя представляет собой интервал времени от момента поступления запроса в ЦОВ до момента окончания обслуживания последнего из всех подзапросов, составляющих этот запрос. Если ресурсов хотя бы одного поставщика услуг, задействованного в обслуживании этого запроса, не хватает для обслуживания соответствующего подзапроса, происходит блокировка запроса пользователя.

Имитационная модель ЦОВ состоит из следующих модулей: «Генератор запросов», модуль «Планировщик», K модулей «Поставщик услуг», модуль «Сборщик», модуль «Сбор статистики» (рис. 1).

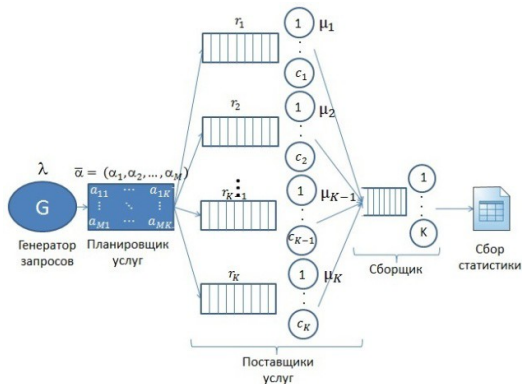


Рис. 1. Структура имитационной модели

В модуле «Генератор запросов» происходит генерация потока запросов в соответствии с экспоненциальным распределением с параметром λ . Сгенерированный запрос, представляющий собой группу заявок, поступает в модуль «Узел-планировщик», который моделируется с помощью СМО типа $M | M | \infty$. Каждой заявке присваивается маркер принадлежности к группе.

В модуле «Планировщик» с вероятностью α_t запрос (группа заявок) объявляется запросом t -типа, тогда в соответствии с матрицей A определяется набор модулей «Поставщик услуг», в которые поступят заявки из этой группы, по одной в каждый модуль. Матрица A имеет K столбцов и M строк, $a_{tt} = 1$, если поставщик задействован в предоставлении t -услуги, $a_{tt} = 0$ в противном случае. Заметим, что $\sum_{t=1}^M a_{tt}$ соответствует длине группы заявок t -типа.

Модуль «Поставщик услуг» номер k представляет собой СМО типа $M | M | \phi | r_k$, $k = 1, \dots, K$. Длительность обслуживания заявки имеет экспоненциальное распределение с параметром μ_k , $0 < \mu_k < \infty$, $k = 1, \dots, K$, причем μ_k не зависят друг от друга. По завершении обслуживания в модуле «Поставщик услуг» заявка переходит в модуль «Сборщик».

В модуле «Сборщик» заявки одной группы ожидают окончания обслуживания последней заявки этой группы. В момент, когда последняя заявка группы поступает в модуль «Узел-сборщик», запрос считается обслуженным.

Модуль «Сбор статистики» позволяет собрать следующие характеристики функционирования модели: время пребывания запроса в системе, время ожидания и время пребывания заявки в модулях «Поставщик услуг», а также вычислить вероятности блокировок запросов для каждого типа услуг, среднюю длину очереди, среднее число занятых приборов, среднее время ожидания начала обслуживания и среднее время пребывания заявки для каждого поставщика услуг.

Исходными данными для имитационной модели являются число K поставщиков услуг, число M типов услуг, интенсивность λ потока запросов пользователей, вектор $\alpha = (\alpha_1, \dots, \alpha_M)$ — выборка типа услуги, матрица A , для каждого типа услуги определяющая набор поставщиков, векторы $c = (c_1, \dots, c_K)$ и $r = (r_1, \dots, r_K)$ — числа приборов и мест в очереди для каждого поставщика услуг. За время моделирования в системе было сгенерировано 10^6 запросов. Сбор статистики производился, начиная с запроса 10^4 .

Литература

1. Borko Furht, Armando Escalante "Handbook of Cloud Computing", Springer New York Dordrecht Heidelberg London, 2010, 634 p.
2. Mohamed Firdhous, Osman Ghazali, Suhaidi Hassan Modeling of Cloud System using Erlang Formulas // 17th Asia-Pacific Conference on Communications (APCC), Sabah, Malaysia 2011.
3. Мокров Е.В., Самуйлов К.Е. Модель системы облачных вычислений в виде системы массового обслуживания с несколькими очередями и с групповым поступлением заявок // Т-Comm - Телекоммуникации и Транспорт. - 2013. - №7. – Принято в печать.

CLOUD COMPUTING MODEL AS A QUEUEING SYSTEM WITH BATCH INPUT

Zakirova R.I., Mikhalkin D.V.

Peoples' Friendship University of Russia, renatazakirova@gmail.com, dvmixalkin@mail.ru

A simulation model of cloud computing center using several of type $M|M|c|r$ is built.

Key words: response time, block probability, CCS.